
SIGILLVM · VNIVERSITATIS · CALIFORNIENSIS
FIAT LVX
MDCCCLXVIII
EX LIBRIS

A TEXT-BOOK OF

PHYSICS

A TEXT-BOOK OF PHYSICS

A. WILMER DUFF, *Editor*

MECHANICS. By A. WILMER DUFF, D.Sc., Professor of Physics, Worcester Polytechnic Institute, Worcester, Massachusetts.

HEAT. By KARL E. GUTHE, Ph.D., Professor and Head of the Department of Physics, University of Iowa, Iowa City, Iowa.

SOUND. By WILLIAM HALLOCK, Ph.D., Professor of Physics and Dean of the Faculty of Pure Science, Columbia University, New York.

WAVE MOTION AND LIGHT. By E. PERCIVAL LEWIS, Ph.D., Professor of Physics, University of California, Berkeley, California.

ELECTRICITY AND MAGNETISM. By ARTHUR W. GOODSPEED, Ph.D., Professor of Physics, University of Pennsylvania, Philadelphia.

ELECTRO MAGNETIC INDUCTION. By ALBERT P. CARMAN, D.Sc., Professor of Physics, University of Illinois, Urbana, Illinois.

CONDUCTION OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY. By R. K. McCLUNG, D.Sc., Late Senior Demonstrator in Physics, McGill University, Montreal; Professor of Physics, Mount Allison University, Sackville, New Brunswick.

BLAKISTON'S SCIENCE SERIES

A TEXT-BOOK

OF

PHYSICS

EDITED BY

A. WILMER DUFF

CONTRIBUTORS

A. WILMER DUFF
E. PERCIVAL LEWIS
KARL E. GUTHE
ARTHUR W. GOODSPEED
WILLIAM HALLOCK
ALBERT P. CARMAN
R. K. McCLUNG

Second Edition Revised
With 525 Illustrations

PHILADELPHIA
P. BLAKISTON'S SON & CO.
1012 WALNUT STREET
1909

Press of
The New Era Printing Company
Lancaster, Pa.

PREFACE TO SECOND EDITION

In preparing a new edition of this text-book the authors have taken advantage of numerous suggestions received from those who have used it in their classes. The part on wave motion has been entirely re-written and numerous changes have been made in several other parts. The distribution of material between large and small print has been changed in various places and it is believed that the part of the book in large print may now be used consecutively and independently as an abbreviated course.

The authors will welcome suggestions for further improvements.

THE EDITOR.

WORCESTER, MASS.,
August, 1909.

EXTRACTS FROM PREFACE TO FIRST EDITION

The preparation of a work of this grade by the collaboration of several writers is a somewhat novel undertaking, and some explanation of its genesis will not be out of place. It represents the attempt of seven experienced teachers of college physics to prepare a text-book that would be more satisfactory to all of them than any existing one. It was, of course, hoped that such a book would also prove acceptable to other teachers. It seemed to the writers that there was a need, and there would be a place, for a work prepared in this way.

One or two remarks as to the character of the book may be permitted. It will in general be found that the writers, while aiming first of all at clearness and accuracy, have preferred terseness to diffuseness. Repetition and amplification are desirable in a lecture. In a printed statement, which may be reread and weighed until mastered, they often discourage thought; and a teacher of Physics might well begin his instruction with the words of Demosthenes, "In the name of the gods I beg you to think." The writers have endeavored to present their subjects simply and directly, avoiding, on the one hand, explanations obvious to any student of fair capacity, and, on the other hand, subtle distinctions and discussions suited to more advanced courses. Some may find the material included in the book too extensive for a single course. If so, a briefer course can be arranged by omitting all of the parts in small print together with as much of those in large print as may seem desirable. There may seem to be some duplication of topics in the work of two contributors. In such cases (which are very few), it will be found that the treatment is from different points of view, appropriate to the respective subdivisions of the subject.

THE EDITOR.

WORCESTER, MASS.,
August 24, 1908.

CONTENTS.

MECHANICS AND PROPERTIES OF MATTER.

HEAT.

TEXT BOOK OF PHYSICS.

MECHANICS AND THE PROPERTIES OF MATTER.

By A. Wilmer Duff, D.Sc.

Professor of Physics in the Worcester Polytechnic Institute, Worcester, Mass.

INTRODUCTION.

1. Physics as a Science. From the evidence of our senses we infer the existence of a great variety of bodies in the physical universe around us. By the use of our senses we also learn that these bodies have various characteristics in common, such as inertia, weight, and elasticity, and these we attribute to the ***matter*** of which in various forms all bodies seem to consist. Matter in itself is inert; the mutual actions of bodies and the effects which they produce on our senses are due to the presence in them of something which is not matter and which is called ***energy***. We shall define the word energy later; the thing denoted by it is known to all as the means which are supplied by the sun, fuels, and elevated bodies of water, and which are required for various familiar operations in nature and industry.

Physics is the Science of the Properties of Matter and Energy. This general description of Physics does not sharply distinguish it from Chemistry and in fact no definite dividing line can be drawn between the two sciences, although in a general way it may be stated that chemistry deals chiefly with questions regarding the composition and decomposition of substances. The different branches of Engineering also treat of the properties of matter, but from the point of view of their useful applications.

A **science** is more than a large amount of information on some subject. In very early times men must have had much valuable

information regarding the physical results of various actions and processes; but it was only when attempts were made to systematize and arrange this knowledge and to seek the relations between the different facts that the science of Physics began. The description of the phenomena of the physical world became more and more scientific as more numerous connections between physical phenomena were discovered and described. At the present time Physics has progressed farther in this direction than any other science, and, in seeking to give a brief account of the present state of the science of Physics, it must be our aim not only to state the most important observed facts but also to show the relations and interdependence of these facts.

It will be seen as we proceed that in some parts of the subject the relations between observed facts are better understood than in other parts. Thus in Mechanics the relations between phenomena have been so well ascertained that we are able to start from a few simple laws regarding the motions of bodies and from these deduce explanations of the most complicated motions. In other parts of the subject we must be content to take from time to time some one principle and trace the logical consequences of it as far as we can, and then proceed to do the same with other principles.

After classifying and studying a group of facts, the process by which we arrive at some underlying principle is called **Induction.** Thus the principle of gravitation was discovered by Newton after a careful comparison of the motions of falling bodies and of the moon and the planets. Having found a general principle underlying and binding together many phenomena, we may reason forward from it and deduce other known or unknown facts, as in Geometry we deduce one proposition from another. This process is called **Deduction.** In a brief account of Physics we must necessarily use deductive more frequently than inductive methods; but, where space will permit, the effort may be made to show how by induction important fundamental principles have been discovered.

2. Measurement. The first condition for success in tracing the connection between the facts in any science is that these facts shall be ascertained as accurately as possible. A qualitative statement of the size or weight of a body, to the effect that it is

large or small, is of very little use. A quantitative description of the same consists in giving the ratio of its size or weight to that of some accepted standard. Such a standard is called a **unit**, and the numerical ratio of the thing measured to the unit is called the **numerical measure** (or numeric) of the thing measured.

Some measurements are *direct*, that is they are made by comparing the quantity to be measured directly with the unit of that kind, as when we find the length of a rod by placing a yard or meter scale beside it. But most measurements are *indirect*. For example, to measure the velocity of a train we measure the distance it travels and the time required, and by calculation we find the number of units of velocity in the velocity of the train.

3. Observation and Experiment. In some branches of science mere *observation*, that is, taking note of circumstances and events, is the chief or only way of obtaining knowledge. For example, the astronomer cannot modify the motions of the heavenly bodies; he must be content to observe. Observation also plays an important part in Physics, but *experiment*, which consists in modifying circumstances or events with a view to making more valuable observations, plays a more important part. Thus if we desire to know how the earth attracts a body and whether the attraction is different at different places, we cannot make much progress if we must confine ourselves to observing bodies falling freely from various heights; but if we modify the fall by attaching the body to a cord and swinging it as a pendulum, we are able to make much more accurate observations and to arrive at valuable information that we could probably never gain by observing free falling bodies. For this reason Physics is chiefly an experimental science, that is to say, the physicist relies on carefully planned experiments to find information and then by methods of reasoning, and especially the condensed accurate form of reasoning called Mathematics, he extracts from the results of the experiment all the information possible.

4. Hypotheses. An event or phenomenon remains obscure or unexplained when its logical connection with other events or phenomena has not been traced. But it is *explained* when it is shown to be connected with other familiar phenomena and the nature of the connection is made clear. Thus the rising of mercury in an

exhausted tube was obscure and unexplained until it was found to be different at different heights along a mountain side and to be connected with the pressure of the air on the mercury in the pool in which the tube stands. The explanation in such a case consists in tracing out the relation of cause and effect between the thing explained and other things. The latter may themselves be still unexplained. Thus the way in which air exercises pressure has only comparatively recently been explained.

A suggested explanation while its correctness is still in doubt is called an **hypothesis.** The hypothesis suggested to account for the pressure of air (or any gas) is that air consists of flying particles which by their bombardment of a surface produce what we call the pressure on the surface; this suggested explanation is called the kinetic hypothesis of gases. The formation of an hypothesis plays a very important part in science, for it stimulates research to test its truth; and even if this particular hypothesis turn out inadequate, in testing it many new facts are usually ascertained and the way paved for arriving at the right explanation. The word **theory** is sometimes used in the same sense as hypothesis, but it is better to restrict it to meaning the extended discussion of an explanation or verified hypothesis. We shall use it in this sense later when speaking of the Kinetic Theory of Gases (§ *227*).

5. Cause and Effect.[1] When a certain event is always followed by a certain other event we are accustomed, in ordinary language, to speak of the former as the **cause** of the latter, and of the latter as the **effect** of the former. Thus the explosion of powder in a gun is spoken of as the cause of the projection of the bullet, and the latter event is described as the effect of the explosion. In speaking of the relation of two things as that of cause and effect we do not merely mean that one has always been observed to follow the other, but we suppose that there is something invariable in the connection between them, that is, we imply our belief that nature will always act in the same way when the circumstances are the same. The principle thus stated is often called that of the **Uniformity of Nature.** There are, however, two circumstances

[1] There is here no attempt to use terms in a critical philosophical sense. The use of such words is unavoidable in an elementary work without confusing circumlocution, and they must be used here in their ordinary sense.

which must be considered as of no importance as regards the connection between causes and effects. These are *time* and *place*. The time of an event is, of course, never repeated, and nothing, so far as we know, ever comes again to exactly the same place, since the sun and all the planets are moving rapidly through space. But if a certain set of circumstances is followed by a certain event, and if this same set of circumstances is repeated, except as regards time and place, the same event will follow.

6. Physical Laws. A careful study of any phenomenon usually enables us to state in a general way what will happen in certain circumstances. Very ancient observation led to the conclusion that bodies when unsupported fall toward the earth. Such a generalization is a *physical law*. A still wider study usually leads to a more general law. Thus the study of falling bodies and of the motion of the moon and of the planets led Newton to the conclusion that each of two bodies is attracted towards the other. The aim of physical research is to obtain physical laws of increasing width and generality. Any such law is very imperfect until it can be stated in exact mathematical form, and this requires careful measurement. By measurement and calculation Newton arrived at the law of attraction between bodies called the Law of Universal Gravitation. Thus a physical law is simply a statement that, given a certain set of circumstances, certain events will follow or it is a statement of some aspect of the Uniformity of Nature.

7. Subdivisions of Physics. The Science of Physics may, for convenience, be divided into the following parts:

1. Mechanics.	**3. Sound.**	**5. Electricity and Magnetism.**
2. Heat.	**4. Light.**	**6. Radioactivity.**

The subject-matter of each of these parts will be described when that part is taken up.

MECHANICS.

8. Mechanics is the branch of Physics which treats of the motions of bodies and the causes of changes in these motions. It is divided into two parts, one, called **Kinematics**, in which the various kinds of motion are described and studied, and the other, called **Dynamics**, in which the causes of change of motion are

studied. Kinematics or the study of motion differs from Geometry in having to consider the element of time. Dynamics is usually divided into two parts, **Kinetics** and **Statics**, the former dealing with bodies in motion and the latter with bodies which, though acted on by causes that tend to produce motion, remain at rest owing to the fact that these influences counteract each other. (Some authors use the term Dynamics in the sense here assigned to Kinetics.) In the following elementary treatment of Mechanics it will not be convenient to treat the various parts of the subject quite separately; each will be taken up in turn as convenience and simplicity may seem to dictate.

KINEMATICS.

THE GEOMETRY OF DISPLACEMENTS.

9. Translation and Rotation. Motions may be divided into two kinds. A moving body has a motion of **translation** when every straight line in the body remains parallel to its original position. Thus a train moving on a straight track and a sled moving down a uniform incline have motions of translation. In such a case all points in the body move in exactly the same way. Hence the motion of the body is completely described when the motion of any point in the body is given, and we may, therefore, in describing the motion of the body, treat it as a single particle or as a point.

A body has a motion of **rotation** when all points in the body travel in circles the centers of which lie in a straight line; the line is called the axis of rotation. This is the motion of a grindstone, a flywheel, or a swing. Any two points in such a body are at any moment moving differently (unless they lie in a plane through the axis and are equidistant from the axis); points farther from the axis move in larger circles and more rapidly than those nearer to the axis.

Many forms of motion are highly complex, but they may in all cases be considered as made up of translations and rotations.

Since the motion of a body which has translation without rotation is the same as that of a point, it is convenient to begin with a study of the motion of a point.

10. Position of a Point. The position of a point is fixed by its

distances, or distances and directions from other points, lines, or surfaces. The simplest way of stating the position of a point is by giving its distance and direction from some other point which we may call the starting point or *origin*.

When we confine our attention to points in a certain line, straight or curved, their positions may be assigned by giving the distance of each point from some assumed origin in that line. One direction away from the origin is taken as positive and the opposite direction as negative. For example the position of any station on a railway line may be fixed by its distance, positive or negative, from some other station taken as origin.

When we confine our attention to points on a surface, plane or curved, the position of each point may be assigned by its distance and direction from some origin on the surface, or, what comes to the same thing, by its distance from each of two lines at right angles passing through the origin. For example, a point on the surface of the earth is described as being a certain distance east or west and a certain distance north or south from the origin.

For points not confined to any line or surface the position of each may be assigned by its distance and direction from some assumed origin in space, or, what comes to the same thing, its distances, positive or negative, from each of three planes intersecting at right angles at the origin.

In the first case position is assigned by one number, in the second by two and in the third by three. A point is said to have one **degree of freedom** when its motion is confined to a definite line, two degrees of freedom when it is confined to a definite surface and three degrees of freedom when it is not restricted in any way.

The above statements of position are statements of *relative position,* that is, statements of the relation of the position of a point to that of some other point taken as origin. Absolute position, or the position of a point without any reference, stated or implied, to any other point or framework of lines, could not be described and no definite meaning could be attached to it. In what follows the word position will always mean relative position, and, unless otherwise stated or implied, the point of reference will be some point on the surface of the earth.

11. Displacements. A change of position is called a **displacement.** In describing a displacement we do not need to make any reference to the time in which the point moves from one position to the other. A description of a displacement consists in a statement of the length and direction of the straight line drawn from the first position of the point to its second position. Thus when a point has moved from A to B it has received a displacement the magnitude of which is the length of the straight line AB and the direction of which is the direction of AB. This displacement we may denote by the symbol $\overrightarrow{AB}$ or $\bar{AB}$, the arrow or stroke being placed above AB to indicate that we are referring not merely to the length of the line AB, but also to its direction from A to B.

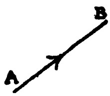


FIG. 1. A displacement is represented by a directed line.

12. Units of Length. To measure or specify a displacement we must use some unit of length. The unit chiefly employed in Physics is the *meter* or one of its multiples or submultiples. The meter is defined as the distance between two lines on a bar of platinum-iridium kept at the International Bureau of Weights and Measures near Paris, when the temperature of the bar is that of melting ice. It was intended by the designers that this length should be one ten-millionth of the distance from a pole of the earth to the equator. One hundredth of the meter is called the centimeter (.01 m.), and this is the unit of length which we shall usually employ. Other decimal fractions of the meter are the decimeter (.1 m.) and the millimeter (.001 m.). For great distances the kilometer (1000 m.) is employed.

The unit of length popularly used in English-speaking countries is the *yard* or one of its well-known multiples or submultiples. The British yard is defined legally as the distance between two lines on a bronze bar kept at the office of the Exchequer in London. The legal definition of the yard in the United States is $\frac{3600}{3937}$ of a meter (see Vol. I of the *Bulletin of the Bureau of Standards,* Washington, D. C.).

13. The Addition of Displacements. If the point that moved from A to B did not travel by the straight line AB but passed

through points C and D, its final displacement was the same as if it had gone by the straight line AB; but the final displacement was the sum of a number of separate displacements, $\overline{AC}$, $\overline{CD}$, $\overline{DB}$. Thus $\overline{AB}$ is the *resultant* or sum of $\overline{AC}$, $\overline{CD}$, $\overline{DB}$, or we may say that by adding $\overline{AC}$, $\overline{CD}$, $\overline{DB}$ we get $\overline{AB}$, or briefly, $\overline{AB} = \overline{AC} + \overline{CD} + \overline{DB}$; but it must be carefully noted that the addition indicated by the sign $+$ is a *geometrical* process, performed by placing the displacements end to end as the sides of a polygon and taking as the sum the displacement from the initial position to the final position.

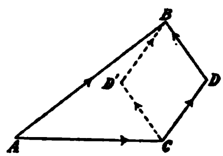


FIG. 2. Geometrical addition of displacement.

If from C we draw a line CD' equal and parallel to DB, and from D' a line $D'B$ equal and parallel to CD, we shall have another path leading from A to B. The displacements $\overline{AC}$, $\overline{CD'}$, $\overline{D'B}$ added together give the same sum as the displacements $\overline{AC}$, $\overline{CD}$, $\overline{DB}$ added together, and for each step in one series there is an equal and parallel step in the other series. It is evident that, so far as addition of displacements is concerned, we may regard $\overline{CD'}$ and $\overline{DB}$ as the same displacement and $\overline{D'B}$ and $\overline{CD}$ as the same displacement. This is consistent with the definition of a displacement as a change of position; for, when a point goes from C to D, it has received the same *change* of position as another point has received when it has gone from D' to B, CD and $D'B$ being equal and parallel. Thus *all displacements which have the same magnitude and direction are equal.*

When two displacements are to be added, the addition may be performed by drawing a triangle. Thus to add $\overline{AB}$ and $\overline{BC}$ we complete the triangle ABC and the sum is $\overline{AC}$. This is called the **triangle method** of adding two displacements. Another method of performing the addition is to construct a parallelogram. If AD be drawn from A equal and parallel to BC, the displacement $\overline{AD}$ is the same as the displacement $\overline{BC}$ and the sum of $\overline{AB}$ and $\overline{AD}$ is $\overline{AC}$, where AC is the diagonal of the parallelogram of which AB and AD are adjacent sides drawn away from A. This is called the **parallelogram method** of adding two displacements. When

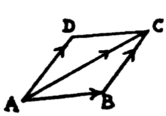


FIG. 3.

several displacements are to be added the addition is performed by constructing a **polygon** as in Fig. 2.

14. Resolution and Subtraction of Displacements. As we may replace any number of displacements by their geometrical sum or resultant, so we may replace a displacement by any number of displacements which added together give the original displacement. This is called ***resolving a displacement into components.*** Thus to resolve a displacement $\overline{AC}$ (Fig. 3) into two components in given directions we draw from A lines in the given direction and then complete the parallelogram $ABCD$ on the diagonal AC; $\overline{AB}$ and $\overline{AD}$ are the components desired, since their sum is $\overline{AC}$.

We may now define the meaning of the ***subtraction*** of one displacement from another. When in ordinary arithmetic we wish to subtract 4 from 10 we resolve the 10 into two parts, 6 and 4, one of which is the same as the number to be subtracted; the other part, 6, is the result of subtracting the 4 from the 10. Similarly, to subtract one displacement from another, we must resolve the latter into two components, one of which is the same as the displacement to be subtracted; the other component is the remainder. Thus to subtract a displacement $\overline{PQ}$ from a displacement $\overline{PR}$ we draw a parallelogram $PQRS$ of which PQ is one side and PR is a diagonal; $\overline{PR}$ is equal to the sum of $\overline{PQ}$ and $\overline{PS}$; subtracting $\overline{PQ}$ from $\overline{PR}$ gives us $\overline{PS}$ as the remainder. Hence, if we use the minus sign to denote the subtraction of one displacement from another, $\overline{PR} - \overline{PQ} = \overline{PS}$.

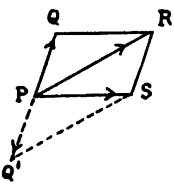


FIG. 4. Subtraction of a displacement.

From the above we may deduce another method of performing the subtraction of a displacement. If PQ' be drawn opposite and equal to PQ, PS is the diagonal of a parallelogram $PQ'SR$ of which PR and PQ' are sides. Thus to subtract $\overline{PQ}$ from $\overline{PR}$ we may add to $\overline{PR}$ a displacement $\overline{PQ'}$ equal and opposite to $\overline{PQ}$. Hence

$$\begin{aligned}\overline{PS} &= \overline{PR} - \overline{PQ}\\ &= \overline{PR} + \overline{PQ'}\end{aligned}$$

From this another conclusion may be drawn. Since PQ' and QP are equal and parallel and in the same direction, $\overline{PQ'} = \overline{QP}$.

Hence it follows from the above equations that $-\overline{PQ}=\overline{QP}$, or the minus sign placed before a displacement reverses its direction without affecting its magnitude.

15. Vector Quantities and Vector Diagrams. Displacements belong to the class of quantities called **vector quantities**, that is, quantities which have *magnitude* and *direction*. Other vector quantities are velocities, forces, etc. The figures in the preceding sections are diagrams of displacements, that is, they are made up of lines representing the actual displacements in magnitude and direction. Thus the diagram might be regarded as a reduced or enlarged picture of the actual displacements. Other vector quantities, *e. g.*, a number of forces, may be similarly represented by a vector diagram by drawing lines each of which stands in magnitude and direction for one of the forces. The lines in such a diagram are called **vectors.** The lengths of any two vectors in such a diagram are to one another as the magnitudes of the forces represented and the angle between the two vectors is the angle between these two forces. After we have defined the meaning of the resultant of a number of forces, it will be seen that it is represented as to magnitude and direction by the vector in the diagram which is the sum of the vectors that represent the separate forces. Similar remarks apply to diagrams of velocities, accelerations, etc.

Quantities which imply no reference to direction are called **scalar** quantities. Such are mass, volume, etc. Each such quantity is assigned by a number without any idea of direction associated with it, and the addition or subtraction of such quantities is performed in the ordinary arithmetic or algebraic manner.

Velocity.

16. Velocity is rate of change of position or *rate of displacement.* Since a displacement has a definite direction as well as a definite magnitude, a velocity also has a definite direction and a definite magnitude, or velocities are vector quantities. Thus "twenty miles an hour" is not a complete statement of a velocity, since it gives only the magnitude of the velocity and does not specify its direction; but "twenty miles an hour eastwards" is a complete statement of a velocity. For clearness such a phrase as "twenty miles an hour" may be called the statement of a *speed,*

which means the mere magnitude of a velocity or a rate of change of position without reference to the direction of the change.

17. Constant Velocity. The velocity of a point is described as *constant* or *uniform* when the displacements of the point in all equal intervals of time are equal. By equal displacements must be understood displacements equal in both magnitude and direction. Hence when the velocity of a point is constant the point moves in a straight line. The magnitude of a constant velocity is measured by the displacement in each unit of time. Hence, if we denote the magnitude of a constant velocity by v and the displacement in time t by s,

$$s = vt.$$

Unit velocity is the velocity of a point that travels unit distance in unit time, *e. g.*, 1 cm. in 1 sec.

18. Variable Velocity. A point has a *variable velocity* when its displacements in equal times are not equal. The displacements in successive equal intervals of time may differ (1) in *magnitude only*, as when a point moves in a straight line with varying speed, or (2) in *direction only*, as when a point moves in a curve with constant speed, or (3) in *both magnitude and direction*, as when a point moves in a curve with varying speed. We shall begin by considering the first of these cases, that of rectilinear motion.

19. Average and Instantaneous Velocity. In *rectilinear motion with variable velocity* how shall we define the magnitude of the velocity? In this case there are two ways open to us. If we divide the whole distance traversed in a certain interval of time by the length of the interval we get the *average velocity* in that interval. If for example we find the whole time required by a train to move from one station to another on a straight track and divide this into the whole distance, we get the *average velocity* between the two stations. In general, denoting the whole distance by s, the whole time by t, and the average velocity by v, we have $\bar{v} = s/t$. Hence

$$s = \bar{v}t.$$

The magnitude of the average velocity in an interval tells us nothing as to the way in which the velocity varies during the interval. If we need to know the character of the motion more

closely we must divide the whole interval into parts and ascertain the average velocity in each. The smaller these parts, the more nearly does the average velocity in any one part represent the actual velocity at any moment in that part. Let us fix our attention on a certain moment at a time t after the beginning of the whole interval. If we proceeded to find the average velocity in a short interval, say Δt, including that moment and if we took successive decreasing values for Δt and found the average velocity in each of these decreasing values of Δt, we would find that the average velocity would rapidly approach a definite limiting value. This limiting value is the *instantaneous velocity* at the moment t. Stated more briefly, *if Δs is the displacement in a small interval of time Δt following the time t, the instantaneous velocity at the time t is the limiting value approached by $\Delta s/\Delta t$ as Δt approaches zero.* This may also be further abbreviated to the form

$$v = \left[\frac{\Delta s}{\Delta t}\right]_{\Delta t = 0}$$

The above definition is also the definition of the derivative of s considered as a function of t. Hence we may also define velocity as the derivative of distance in the direction of motion with respect to time, or

$$v = \frac{ds}{dt}$$

When the velocity of a point is constant, the instantaneous velocity, as defined above, is the same as the velocity of the point, as defined in § 17. For the values of $\Delta s/\Delta t$ at different moments in any interval t are equal. Hence, if s is the whole distance traversed in the time t, each value of $\Delta s/\Delta t$ is equal to s/t, which is the distance traversed in unit time.

When the instantaneous velocity of a point is variable, we may also state its magnitude in terms of an equal constant velocity. Suppose that when the instantaneous velocity is v, the point begins to move with a constant velocity equal to v. The magnitude of this constant velocity is the distance the point would travel in unit time. Hence we may say that *the instantaneous velocity of a point is equal to the distance the point would travel if its velocity remained constant for a unit of time and equal to the instantaneous velocity.*

20. The Unit of Time. To measure or specify a velocity we must use some *unit of time*. The unit of time usually employed in Physics is the **mean solar second.** This is defined as $\frac{1}{86400}$ of the mean solar day, which is the average throughout a year of the time between two successive transits of the sun across the meridian at any place. It is the second of the ordinary clock or watch when it keeps correct time.

21. Curvilinear Motion. When the displacements of a point in successive equal intervals are in different directions, the point is moving in some curved path. This, for example, is the case when a ball is thrown obliquely upward or when a train is moving on a curved track. If the position of the point at a certain time t is P and at a somewhat later time, say $(t + \Delta t)$, is Q, the displacement in this time is $\bar{PQ}$. If we denote the length of PQ by Δs and consider the limiting value of $\Delta s/\Delta t$ as before, we get the instantaneous velocity of the point at the time t when the point is at P. As PQ is decreased the chord PQ finally approaches without limit to the tangent at P; hence the direction of the instantaneous velocity at P is along the tangent at P. While this is the proper meaning of the rate of displacement at P, we should arrive at the same value for the instantaneous velocity if we took Δs to mean the length of the arc PQ, and supposed it successively diminished by the approach of Q toward P; for the chord and the arc would in the limit have a ratio of unity.

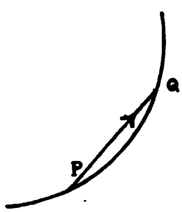


FIG. 5.

22. The Graph of a Varying Velocity. When any quantity is variable, much valuable information can frequently be derived from the properties of a curve drawn to represent the varying quantity. A curve drawn to represent the magnitude of a varying velocity is called a *velocity curve*. Let OA be a straight line of which the length OA stands for the length of the interval, t, in which we wish to consider the motion. Divide OA up into a very large number of small equal parts. At O erect a perpendicular OB to represent the magnitude of the velocity at the beginning of the interval t. Erect similar perpendiculars to represent the instantaneous velocities at the beginnings of the other parts of the inter-

val, and through the upper ends of these perpendiculars draw a smooth curve BC.

Consider one of these short intervals, ab. If the velocity throughout this short interval had been the same as at the beginning of the short interval, say v, the distance traversed in the short interval would have been $v \times ab$ or the unshaded rectangle above ab. If the velocity throughout the short interval had been the same as that at the end of the short interval, say v', the distance would have been $v' \times ab$ or the area of the unshaded rectangle plus that of the small shaded rectangle above it. The real distance in the interval is intermediate between these two. Applying the same reasoning to all the small intervals in succession, we see that the whole distance is something between that represented by the whole unshaded area between BC and OA and that represented by the unshaded area plus the shaded area. If now we suppose the number of the small intervals increased without limit so that each becomes vanishingly small, the shaded area will decrease without limit until it vanishes and the area between the curve BC and the line OA will represent the actual distance in the time t.

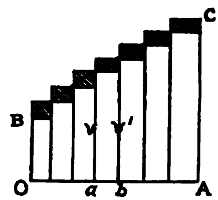


FIG. 6. Graph of a velocity.

To bring out more clearly the meaning to be attached to the word "represent" in the above, let us first suppose that OA contains as many units of length as t contains units of time, and that OB contains as many units of length as the velocity it stands for contains units of velocity. Each unit of area will then stand for a unit of distance traversed by the moving point, and the whole area will contain as many units of area as the distance traversed contains units of length. But if each unit of length along OA stands for m units of time and each unit of length along OB stands for n units of velocity, the whole area will be mn times smaller than it would have been on the first supposition, and, to get the whole distance, we shall have to multiply the whole area by mn.

23. The Resultant of Simultaneous Velocities. A man sitting in a train has the velocity of the train, but when he gets up and moves about he has an additional velocity which may or may not

be in the same direction as the first velocity. Similarly a launch floating down with the current in a river has the velocity of the current; but, if it has a propeller in motion, it has another velocity in addition to the first. *When a body has two or more simultaneous velocities, it pursues some definite path and its velocity in the path is called the* **resultant** *of the simultaneous velocities.*

From this definition of the resultant of any number of simultaneous velocities it can be shown that the magnitude and direction of the resultant velocity can be deduced from the separate velocities by the triangle, parallelogram, or polygon method of adding vectors. Consider first the case of two constant velocities and draw a diagram, in which $\overline{AB}$ and $\overline{AC}$ stand for the two velocities. Complete the parallelogram $ABCD$. We shall show that $\overline{AD}$ stands for the resultant velocity. Since the velocities are constant $\overline{AB}$ and $\overline{AC}$ represent in magnitude and direction the component displacements in unit time, and the sum of these displacements is represented by $\overline{AD}$ (§ 13), which, therefore, represents *the resultant displacement in unit time.* Hence AD represents the resultant velocity. Thus the parallelogram method applies to the addition of constant velocities, and the same must be true of the other methods, which are essentially the same.

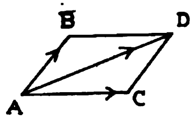


FIG. 7.

When the component velocities are not constant we can add their instantaneous values by the vector methods referred to. The proof of this statement is the same as above except that $\overline{AB}$, $\overline{AC}$, and $\overline{AD}$ will now stand for the displacements that would take place in unit time if the velocities remained constant that long.

It is readily deduced from the above that the *difference* of two velocities $\overline{AB}$ and $\overline{AC}$ is a velocity $\overline{CB}$. For from the triangle ACB it is seen that a velocity $\overline{CB}$ added to the velocity $\overline{AC}$ will give a velocity $\overline{AB}$. Also since $\overline{CD}$ represents the same velocity as $\overline{AB}$, $\overline{CB}$ is the difference of velocities $\overline{CD}$ and $\overline{AC}$. Briefly stated

$$\begin{aligned} \overline{AC} + \overline{CB} &= \overline{AB} \\ \therefore \overline{CB} &= \overline{AB} - \overline{AC} \\ &= \overline{CD} - \overline{AC} \end{aligned}$$

24. Formula for Resultant. Let v_1 and v_2 be the respective magnitudes of two component velocities of a moving point, and let these velocities be represented by $\overline{OA}$ and $\overline{OB}$ respectively. Also let V be the magnitude of the resultant velocity, which is represented by $\overline{OC}$, where OC is the diagonal of the parallelogram of which OA and OB are sides. By a well-known trigonometrical formula

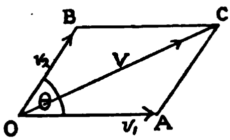


FIG. 8.

$$OC^2 = OA^2 + AC^2 - 2OA \cdot AC \cos OAC$$

Denote the angle AOB, which is the angle between the direction of the two components, by θ. Then the angle OAC equals $(180° - \theta)$ and therefore $\cos OAC = -\cos\theta$. Since OA, OB, and OC are proportional to v_1, v_2, and V respectively,

$$V^2 = v_1^2 + v_2^2 + 2v_1 \cdot v_2 \cos\theta$$

By this formula we can calculate the magnitude of V when v_1, v_2, and θ are known.

When $\theta = 0$, that is, when the components are in the same direction, $\cos\theta = 1$ and the formula for V gives $V = (v_1 + v_2)$. When $\theta = 180°$, that is, when the components are in opposite directions,

$$\cos\theta = -1 \text{ and } V = \pm(v_1 - v_2).$$

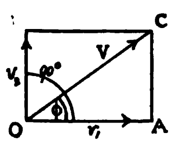


FIG. 9.

If $\theta = 90°$, that is, if the components are at right angles, $\cos\theta = 0$ and

$$V^2 = v_1^2 + v_2^2$$

and if ϕ be used to denote the angle AOC which the resultant makes with the component of magnitude v_1,

$$\tan\phi = \frac{AC}{OA} = \frac{v_2}{v_1}$$

25. Resolution of a Velocity into Components. Since two velocities taken together are equivalent to a single velocity called their resultant, we may reverse the process and suppose any velocity replaced by any two velocities which added are equivalent to the original velocity. This is called *resolving a velocity into com-*

ponents. To thus resolve a velocity we must draw a parallelogram of which the diagonal stands for the velocity to be resolved. Now any number of parallelograms can be drawn with a given line as diagonal; but, if the directions of the sides are specified, only one solution is possible. Hence to resolve a given velocity into components in two given directions is a definite problem which may be solved graphically by constructing a parallelogram.

The most important case of the above is when the directions of the components are at right angles. Thus if the velocity is V in the direction Oc and if $\overline{OC}$ is taken to represent V and if Oa and Ob are to be the directions of the components, we draw from C perpendiculars, CA and CB, to Oa and Ob respectively. Then $\overline{OB}$ and $\overline{OA}$ are the desired components in the specified directions.

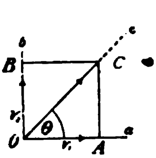


FIG. 10.

If the direction Oa makes an angle θ with the direction of V and if we denote the components in the directions Oa and Ob respectively by v_1 and v_2,

$$v_1 = V \cos \theta; \ v_2 = V \sin \theta.$$

It should be noted that θ stands for an angle that may be either positive or negative. We may regard θ as the angle through which a line, starting from the position Oa, must revolve about O to reach the position Oc; and when the revolution is counter-clockwise, it is customary to regard such an angle as positive, the opposite direction of revolution corresponding to a negative angle. If we make this agreement as regards the sign of θ we must keep to it as regards the right angle that Ob makes with Oa, that is to say the right angle and the angle θ must be measured away from Oa in the same direction, namely counter-clockwise.

Acceleration.

26. Acceleration is rate of change of velocity. A change of velocity has a definite direction as well as a definite magnitude. Hence acceleration is a quantity which has both direction and magnitude, that is, acceleration is a vector quantity.

An acceleration may be either constant or variable. The acceleration of a point is *constant* when the velocity of the point

changes by equal amounts in equal intervals of time. By equal changes of velocity must be understood changes of velocity that are equal in magnitude and in the same direction. When the changes of velocity in equal intervals of time are not equal, the acceleration is *variable*.

The statement that the velocity of a point is variable may refer to a change in the magnitude of the velocity, to a change in the direction of the velocity, or to a change in both. Hence we shall have three cases of acceleration to consider, (1) the acceleration of a point when the velocity of the point is constant in direction but variable in magnitude, (2) the acceleration of a point when the velocity of the point is constant in magnitude but variable in direction, (3) the acceleration of a point when the velocity of the point is variable in both magnitude and direction.

The simplest case is when the velocity of the moving point is constant in direction and when the acceleration is constant and in the direction of the line of motion. This is illustrated by a body dropped from a height and falling in a straight line.

The *magnitude of a constant acceleration* is the magnitude of the velocity added in each unit of time, and the direction of the acceleration is the direction of the added velocity. The *unit of acceleration* is that of a point the velocity of which increases by unit velocity in unit time. When the cm. is taken as unit of length and the sec. as unit of time, the unit of acceleration is such that the velocity increases by one cm. per sec. in each second, or, briefly, one cm. per sec. per sec.

27. Motion in a Straight Line with Constant Acceleration. In considering the motion of a point along a straight line we take one direction along the line as positive and the opposite direction as negative. Let u be the velocity of the point at the beginning of an interval of time of length t, and let v be its velocity at the end of the interval. The increase of velocity is $(v-u)$ and the increase per unit time is $(v-u)/t$. This is, therefore, the magnitude of the constant acceleration which we shall denote by a. Hence

$$v = u + at \qquad (1)$$

This very important equation is simply a statement that the final velocity (at the end of the time t) is equal to the initial

velocity (at the beginning of t) plus the increase of velocity, and the increase of velocity is equal to the acceleration multiplied by the time.

To find how far the point travels in the time t let us consider the form of the velocity curve (§ 22) in the present case. The changes of velocity in equal short intervals of time are equal. Hence in Fig. 6 the differences between each ordinate and the next in order are equal, and the velocity curve is therefore a straight line as in Fig. 11. Draw BD parallel to OA. The whole area $OBCA$ consists of two parts, that of the rectangle $OBDA$ and that of the triangle BDC. OB represents the initial velocity u and we shall suppose that the figure is drawn to such a scale that OB contains as many units of length as u contains units of velocity and that the same is true of AC which represents the final velocity v. The height of the triangle DC represents in the same way the increase of velocity at. OA represents the time t and we shall suppose that OA contains the same number of units of length as t contains units of time. The whole area is therefore $(ut + \frac{1}{2}t \cdot at)$. Hence if s is the whole distance traversed in the time t,

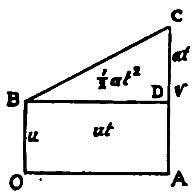


FIG. 11.

$$s = ut + \tfrac{1}{2}at^2 \tag{2}$$

This very important equation consists of two parts. The part ut is the distance the point would have travelled in the time t if its velocity throughout t had remained constant and equal to the initial velocity u. The part $\frac{1}{2}at^2$ is the additional distance due to the acceleration, that is the distance the point would have gone if it had started from rest with an acceleration a.

Between (1) and (2) we may eliminate t and so find an expression for the final velocity in terms of the initial velocity, the acceleration, and the distance.

$$\begin{aligned} v^2 &= u^2 + 2uat + a^2t^2 \\ &= u^2 + 2a(ut + \tfrac{1}{2}at^2) \\ &= u^2 + 2as \end{aligned} \tag{3}$$

These three equations are of great importance.

Another expression for the area $OBCA$ is $\frac{1}{2}(OB + AC) \cdot OA$.

Hence the distance is also given by the formula

$$s = \frac{v + u}{2} \cdot t$$

From this it follows that the average velocity, which equals the total distance divided by the time (§ 19), is equal to one half of the sum of the initial velocity and the final velocity.

Equation (2) may be readily obtained by means of the Integral Calculus. The distance travelled in a short time dt when the velocity is v is vdt. Hence the whole distance, $s, = \int_0^t vdt = \int_0^t (u + at)dt = ut + \frac{1}{2}at^2$.

28. Galileo's Experiments. The very important relations expressed by (1) and (2) were discovered by Galileo by studying the motion of falling bodies, and this discovery was the beginning of Kinetics. Before that time nothing was known as to the way in which the velocity of a body increases as it falls. Galileo thought the law of increase expressed by (1), namely, that the increase of velocity is proportional to the time, was probably correct; but the instrumental means at his command did not enable him to test it; so he deduced (2), practically by the graphical method given in § 27, and then tested it. To avoid having to deal with any great velocities such as that of a body falling vertically, he tested the rolling of a ball down an inclined plane, assuming that both motions would follow the same law. The result confirmed his formula.

29. Acceleration of Free Fall. We shall assume as an experimental fact, discovered by Galileo, that at any one place all bodies falling freely would have the same acceleration if it were not for the effect of air friction. The latter is very small in the case of dense solids, such as blocks of metal, falling moderate distances, and may usually be neglected. The acceleration of free fall, or the acceleration of gravity, as it is often called, is usually denoted by g. In centimeters and seconds g is about 980, though slightly different at different points on the earth's surface, and in feet and seconds it is about 32.2. Hence, from § 27, when a body is projected vertically downward with a velocity u its velocity and

distance after an interval t may be found from

$$v = u + gt$$
$$s = ut + \tfrac{1}{2} gt^2$$
$$v^2 = u^2 + 2 gs$$

When the direction of projection is upwards we may take upwards as the positive direction and g, being downwards, will then be negative. In this case

$$v = u - gt \qquad (1)$$
$$s = ut - \tfrac{1}{2} gt^2 \qquad (2)$$
$$v^2 = u^2 - 2 gs \qquad (3)$$

At the highest point $v = 0$; hence from (1) we have $t = u/g$. Substituting this in (2) we get for the *height of ascent* $s = \frac{1}{2}u^2/g$. This also follows from (3) by putting $v = 0$. The time of return to the ground is got by putting $s = 0$ in (2). This gives $t = 2u/g$, showing that the whole time of rise and fall equals twice the time of ascent, or that *the time of rise equals the time of fall.* It follows from (3) that the velocity of return to the starting point, that is when s is again zero, equals the velocity of projection in magnitude, but is in the opposite direction. It must, however, be remembered, that these statements are true only for moderate velocities. At high velocities, such as those of a bullet, air-resistance greatly modifies the motion.

The value of g at any station of observation depends on the latitude of the station and also on the height of the station above sea level. The results of very careful experiments show that at a station in latitude λ and at an elevation of l meters above sea level,

$$g = 977.989\,(1 + .0052 \sin^2 \lambda - .0000002\, l).$$

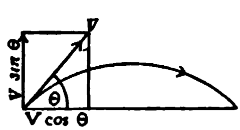


FIG. 12. Path of a projectile.

30. Motion of a Projectile. When a body is thrown obliquely into the air its motion may be considered as consisting of a horizontal part and a vertical part. The vertical part is subject to a constant acceleration g downward; while, since there is no horizontal acceleration (if we may neglect air-friction), the horizontal part of the motion is a constant velocity. If the mag-

nitude of the velocity of projection is V and the direction of projection makes an angle θ with the horizontal, the velocity V may be resolved into a component $V \cos \theta$ in a horizontal direction and a component $V \sin \theta$ in a direction vertically upward. If then x is the horizontal distance traversed in time t,

$$x = Vt \cos \theta \qquad (1)$$

and if, at the time t, the vertical distance attained is y,

$$y = Vt \sin \theta - \tfrac{1}{2} gt^2 \qquad (2)$$

Thus the vertical motion is the same as that of a body thrown vertically upward with a velocity $V \sin \theta$. Hence (§ 29) at time $(V \sin \theta)/g$ the body will have just lost its vertical velocity and will therefore be moving wholly in a horizontal direction; and at that moment the height will be $(V^2 \sin^2 \theta)/2g$. At time $(2V \sin \theta)/g$ the body will have returned to its original level and the distance horizontally from its starting point will then be $V \cos \theta \cdot (2V \sin \theta)/g$ or $(V^2 \sin 2\theta)/g$. Now since $\sin 2\theta$ has its maximum value, unity, when 2θ is 90°, that is, when θ is 45°, it follows that the greatest horizontal range for a given velocity, V, of projection is V^2/g and is obtained by making the angle of projection 45°.

If it be desired to find the constant relation that holds between x and y during the motion, the value of t taken from (1) may be substituted in (2) and we shall get

$$y = x \tan \theta - x^2 \frac{g}{2v^2 \cos^2 \theta}$$

the equation of a parabola referred to axes through the point of projection. Hence the path of the projectile is a parabola.

As in the case of § 29, these results are approximately correct only in the case of the moderate velocities for which air-friction is negligible.

31. Variable Acceleration. When the acceleration of a point is variable we can no longer measure it by the actual increase of velocity in any time. We may, however, divide the magnitude of the increase of velocity in any time by the time and call this the magnitude of the *average acceleration* in that time, the direction of this average acceleration being the direction of the increase of velocity. The *instantaneous value of the acceleration* is defined

much as in the case of instantaneous velocity, namely, as the value to which the average acceleration approaches as the interval is diminished without limit.

It follows from the above that we may also express acceleration as the derivative of v as follows

$$a = \frac{dv}{dt}$$

A variable acceleration may be variable as regards magnitude or direction or both. In the following we shall consider the case of an acceleration that is constant in magnitude but variable in direction.

32. Acceleration of a Point that Moves with Constant Speed in a Circle. Let P be a momentary position of the moving point and let Q and R be points on the circumference equidistant from P, the time of motion from Q to R being t. At Q and R draw tangents that meet in T. QT and TR are equal in length and may therefore be taken to represent the *velocities* at Q and R respectively.

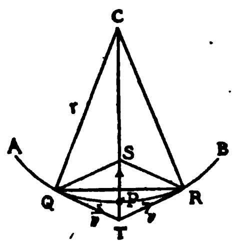


FIG. 13.

Complete the parallelogram $QTRS$. The difference of velocities $\overline{QT}$ and $\overline{TR}$ is a velocity $\overline{TS}$ (§ 23). Thus the moving point suffers a change of velocity represented by $\overline{TS}$ and therefore *directed toward the center C*. If a be the mean acceleration in the time t, the change of velocity will be at. Hence, denoting the constant speed (that is, the constant magnitude of the velocity) by v,

$$\frac{at}{v} = \frac{TS}{QT}$$

It is easily shown that the triangles SQT and QCR are similar. Therefore

$$\frac{TS}{QT} = \frac{QR}{CR} = \frac{\text{chd } QR}{\text{arc } QR} \cdot \frac{\text{arc } QR}{CR}$$

Now the arc QR, being described in time t with speed v, is equal to vt. Hence by substitution and cancellation of t

$$\frac{a}{v} = \frac{\text{chd } QR}{\text{arc } QR} \cdot \frac{v}{r}$$

In this a is the mean acceleration while the point moves from Q to R; but if we now suppose the time t indefinitely short, a will be the acceleration of the moving point when it is at P and the ratio of the chord to the arc (each being indefinitely small) will be unity. Hence

$$a = \frac{v^2}{r}$$

and this acceleration is always directed toward the center C. It is, therefore, an acceleration of constant magnitude but of variable direction.

33. Curvilinear Motion. If in the preceding the speed were not constant, there would, in addition to the acceleration toward the center, be an acceleration along the tangent. The first acceleration would have the effect of changing the direction of the velocity, while the second would have the effect of changing the magnitude of the velocity, that is the speed.

When a point moves in a curve of any form, it may be regarded as moving at any moment in a circle which coincides at that point with the curve; this circle is called the circle of curvature at that point on the curve. From the radius of the circle of curvature at a point on the curve we can calculate the acceleration toward the center of the circle of curvature. If the speed of the point is not constant, the point must also have an acceleration along the tangent to the curve.

34. Addition of Accelerations. A moving point may have two or more accelerations simultaneously. Thus a man at rest on the deck of a ship which is moving with an acceleration has one acceleration, that of the ship. If he moves across the deck with an acceleration independent of the motion of the ship, he has a second acceleration. In any such case the moving body travels in some curve with a definite acceleration which is called the *resultant* of the component accelerations.

We can readily show that the resultant acceleration may be deduced from the component accelerations by the vector method of addition, that is by the construction of a triangle, parallelogram or polygon, the sides of which represent the separate accelerations. For let $\overline{AB}$ and $\overline{AC}$ represent two constant accelerations possessed simultaneously by a point. Since the acceleration represented by $\overline{AB}$ is constant, the change of velocity it produces in unit time is also represented by

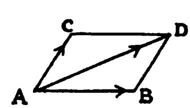


FIG. 14.

$\overline{AB}$. Similarly $\overline{AC}$ represents the change of velocity in unit time due to the second constant acceleration. The resultant change of velocity is found by completing the parallelogram *ABDC;* hence $\overline{AD}$ is the resultant change of velocity in unit time, that is, the resultant acceleration. The same method of reasoning is applicable when the accelerations are variable; the only difference being that $\overline{AB}$ and $\overline{AC}$ and $\overline{AD}$ all represent velocities that would have been added in unit time, if the accelerations remained constant that long.

35. Resolution of an Acceleration into Components. Since two or more accelerations may be replaced by their resultant, it follows that an acceleration may be resolved into two or more components by the ordinary methods. The case in which an acceleration is resolved into two components at right angles is especially important. As an example suppose a body rests on a smooth plane the inclination of which to the horizontal is i. If the body were not supported it would fall with an acceleration of g. The acceleration may be resolved into a component $g \cos i$ perpendicular to the plane and a component $g \sin i$ parallel to the plane. The component perpendicular to the plane has no effect, since motion perpendicular to the plane is prevented, whereas the other component causes it to slide down the plane with an acceleration $g \sin i$. Thus the motion down the plane may be calculated by the formulæ of § 29, a being replaced by $g \sin i$.

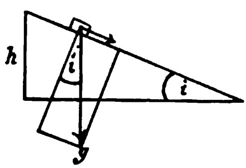


FIG. 15. Acceleration down a plane.

From this we may deduce one result of importance. The velocity after a distance of descent s down the plane is given by

$$\begin{aligned} v^2 &= u^2 + 2as \\ &= u^2 + 2g \sin i \cdot s \end{aligned}$$

and if h is the distance of descent measured vertically $h = s \cdot \sin i$. Hence

$$v^2 = u^2 + 2gh$$

Now this is the formula we should have been led to if we had sought the velocity attained by a free vertical fall through a dis-

tance h. Hence the *speed attained by a body which slides without friction through a certain vertical distance is the same as if the body had fallen that distance vertically.*

DYNAMICS.

Force and Mass.

36. In the preceding we have considered various cases of motion without any reference to the influences that affect the motions of bodies, just as in Geometry we study lines and figures without any reference to particular bodies. We must now consider those relations between bodies on which changes of motion depend.

Isaac Newton was the first who attained clear ideas as to the relations between bodies and their motions. His treatment of the subject was founded on three fundamental principles which he called Axioms or Laws of Motion. These axioms are so simple that they are recognized as very probably true as soon as their meaning is grasped. The proof of their correctness is, however, the fact that all deductions from them are found to be verified by observation and experiment.

37. Newton's First Law of Motion. *Every body persists in its state of rest or of uniform motion in a straight line, unless it be compelled by some force to change that state.*

This law may be divided into two parts, a *statement* and a *definition.* The statement is that any change of velocity of a body, that is any acceleration, is due to some external influence, and a body free from external influences would necessarily have a constant velocity. This was a complete denial of what had been supposed to be true up to the time of Galileo (who died in 1642, the year in which Newton was born); for until then it was supposed that a body free from external influences would come to rest. We cannot, of course, free any body entirely from external influences; but we can greatly diminish these influences and with each diminution the velocity becomes more nearly constant. The most common hindrance to steady motion is friction. A stone given a push along a rough road is quickly stopped by friction; on a smooth floor it will continue longer in motion; a well polished stone started on smooth ice will continue in motion for a great distance. Such considerations make it seem probable

that, if freed from external influences, a body would move with constant velocity; they do not, however, amount to a proof of the statement in the first law of motion. The proof of the law is that all of the innumerable deductions made from it and the other laws of motion are verified by experience.

The law implies a definition of **force**. This is usually given in the form "Force is whatever changes or tends to change the motion of a body" or "Force is that which produces acceleration." Thus friction, the pull of a stretched spring, the attraction of the earth on a body, etc., are forces; when a body revolves in a circle it has an acceleration toward the center and must therefore be acted on by some force. What exerts a force on a body is, of course, some other body. Thus the friction opposing the motion of a vehicle is due to the earth and the pull of a spring is due to the spring, etc. The word force is therefore a name which we give to that influence of one body on another by which the first changes the motion of the second.

The property a body has of tending to persist in its state of motion or of rest is called **Inertia.**

38. The Mass of a Body. Common experience shows that when a given force is applied to a body the magnitude of the acceleration depends on some property of the body. Thus a horizontal spring kept stretched to a definite length, say one foot, will apply a definite force to the body to which it is attached. If attached to a cubic foot of lead supported without friction it will produce a certain acceleration; but the acceleration will be different if a cubic foot of wood be substituted for the lead. The difference is not due to the difference in the weights of the bodies, since weight is a force that acts vertically and does not affect the horizontal motion of the bodies. The difference is due to what we call the masses of the bodies.

To attach a definite meaning to the word **mass** we must define what is meant by the ratio of the masses of two bodies. *The ratio of the masses of two bodies is the inverse of the ratio of the accelerations that a given force imparts to the bodies when applied to them in succession.* For example if a body A acted on by a certain force receives twice the acceleration that a second body B receives when acted on by the same force, the mass of A is half

as great as the mass of B and so for other ratios. Hence if we adopt, as we presently shall, a certain body as a body of unit mass, the mass of any other body becomes definite.

In the above we have defined the ratio of two masses by means of the ratio of the acceleration imparted to them by *some particular force*. This at once suggests the question: Would the ratio be found different if some other force were used in the test? If so the word mass as applied to a body would have no definite meaning except in relation to a particular force. But as a matter of fact *the ratio is found to be the same no matter what may be the force chosen for the test*. This is a very important statement but we do not need to state it as a separate fundamental principle since, as we shall see, it is included in Newton's Second Law of Motion. The fact that *bodies have definite masses*, the same no matter what their accelerations or the forces acting on them, was one of Newton's most important discoveries.

Some persons find difficulty in accepting the above definition of the ratio of two masses, because they cannot see an easy means of applying it directly to comparing masses. It is, however, not given as a practical method of comparing masses; but it leads, as we shall see later, to a very practical method (§ 42).

39. Units of Mass. The unit of mass chiefly employed in Physics is the *gram*, which is defined as one one-thousandth of the mass of a block of platinum kept at Sèvres, near Paris, and known as the *Kilogram prototype*. Fractions and multiples of the gram are named as follows:

Milligram	= .001 g.	Kilogram	= 1000 g.
Centigram	= .01 g.	Hectogram	= 100 g.
Decigram	= .1 g.	Dekagram	= 10 g.

$$\text{Metric ton} \begin{cases} = 1{,}000{,}000 \text{ gms.} \\ = 1000 \text{ Kilograms} \end{cases}$$

In English-speaking countries for commercial and industrial purposes the *pound* is used as unit of mass. It is defined as the mass of a certain block of platinum kept at the Exchequer in London. It is worth remembering that 1 kgm. = 2.20 lbs. approximately and that 1 pound = 454 gms. approximately.

40. Ratio of Forces. Different forces applied to a body give it different accelerations. For example, if a heavy body be hung from the ceiling by a cord and a horizontal cord be attached to it, a

small pull will start it slowly while a stronger pull will start it more rapidly. Or, if a horizontal spiral spring, kept stretched to a definite length, were applied to a body supported with very little friction on a horizontal table, a definite acceleration would be produced. If this experiment were repeated with the spring stretched to a different length, a different acceleration would result. These illustrations would be somewhat difficult to carry out accurately, but they will help to make clear the following definition of the ratio of two forces, and from this we shall be able to deduce a more accurate method of finding the ratio either by calculation (§ 42) or by static experiments (§ 52).

The ratio of two forces is the ratio of the accelerations they can impart to a given body. For definiteness we shall suppose that the body referred to is one of unit mass. If now we take any force as unit force, the magnitude of any other force becomes definite. For simplicity we shall usually take *as unit force that force which, acting on unit mass, gives it unit acceleration.* A force which gives unit mass two units of acceleration will then be a force of two units, and so on.

In the above we have defined the ratio of two forces by the ratio of the accelerations they impart to some particular body. This definition would not be of much value if the ratio obtained depended on the particular body chosen. As a matter of fact the ratio of two forces as defined above is the same, no matter what body is chosen for the test. This statement, while very important, does not need to be stated as a separate fundamental principle since, as we shall see, it is included in Newton's Second Law of Motion.

41. Momentum. Every one is aware that certain properties of moving bodies depend on mass and velocity conjointly. Thus the length of time required by a locomotive to start a train depends on both the mass of the train and the velocity to be imparted to it, and the same is true of stopping it. Hence we find it convenient to define a property depending on mass and velocity conjointly. *Momentum* is defined as *the product of mass and velocity.* Since the velocity has direction as well as magnitude, while the mass is a mere number, the momentum of a body is a vector quantity, the direction of which is that of the velocity.

(What we now call momentum Newton called *quantity of motion* as distinguished from *rate of motion* or velocity.)

When the velocity of a body changes, its momentum also changes. Since the mass of the body is constant, any change in the momentum of a body must be due to a change of its velocity and the change of momentum must equal the product of the mass and the change of velocity. Hence when the momentum of a body is changing the rate of change of momentum equals the product of the mass and the rate of change of velocity, that is the product, ma, of the mass m and its acceleration a.

42. Newton's Second Law of Motion. *The rate of change of the momentum of a body is proportional to the force acting on the body and is in the direction of the force.*

To reduce this statement to a mathematical formula let us suppose that a force F_1 acting on a mass m_1 gives it an acceleration a_1, and that a force F_2 acting on a mass m_2 gives it an acceleration a_2, and so on for any number of forces and masses. Then

$$F_1 : F_2 \cdots :: m_1a_1 : m_2a_2 \cdots$$

or

$$\frac{F_1}{m_1a_1} = \frac{F_2}{m_2a_2} = \cdots = \text{a constant, say } k.$$

From this it is seen that the assumptions stated in §§ 38, 40 are included in the second law. For if any force F act in succession on two bodies of masses m_1 and m_2 respectively we may in the above put F_1 and F_2 both equal to F. Hence $m_1a_1 = m_2a_2$ or $m_1/m_2 = a_2/a_1$ which is the assumption of § 38. Again if two forces F_1 and F_2 act in succession on any mass m we may put m_1 and m_2 both equal to m. Hence $F_1/F_2 = a_1/a_2$ which is the assumption of § 40.

The meaning of the ratio in the above formula for the Second Law should be carefully considered. The numerator is a number, namely, the number of units of force acting on a body. The denominator is also a number, obtained by multiplying the number of units of mass in the mass of the body by the number of units of acceleration imparted to the body. Hence the ratio k is a number and this number is the same for all the ratios. How large the number k will be will depend on the magnitudes of the units employed in measuring F, m, and a. If we start with any convenient units of mass and acceleration and if we then adopt as unit force *that force which acting on unit mass will give it unit acceleration,* it follows that when m and a are each unity F must

be unity and in this case k must be unity. With this definition of unit force the formula for the second law takes the simple form

$$F = ma$$

Since, as Galileo found (and as Newton and Bessel proved more completely), all bodies fall with the same acceleration (allowance being made for air friction), it follows from the above formula that *the masses of bodies are proportional to their weights.* This is the principle of the common balance by which the masses of bodies are compared by comparing their weights.

43. Units of Force. A *unit of force* defined as in the preceding section, namely, *such that when acting on unit mass it gives it unit acceleration,* is called an *absolute unit;* it is the unit of force which bears the simplest possible relation to the units of mass, length and time. When these units are the gm., the cm., and the sec., respectively, we have what is called the C. G. S. system of units, and the absolute unit of force in this system, which is the force that will give to one gm. an acceleration of one cm. per sec. per sec. is called the **dyne**. The absolute unit in the lb. ft. sec. system is called the *poundal.* It is little used.

Since a body of one gm. mass allowed to fall freely has an acceleration of 980 cm. per sec. per sec. (approx.), the force acting on it, which is its *weight,* must be 980 times larger than the unit of force, that is the dyne is about 1/980 of the weight of a gram, and is therefore slightly greater than the weight of a gram, and is, therefore, slightly greater than the weight of 1 mg.

Engineers in English speaking countries use the weight of a pound as unit of force. With this unit of force and the mass of a pound as unit of mass the value of k in the formula (§ 42) for the second law cannot be unity. For when the pound mass, that is unit mass, is allowed to fall it is acted on by the weight of a pound which is unit force, and its acceleration is 32.2. Hence $k = 1/32.2$ and with these units the second law must be written in the form

$$F = \frac{1}{32.2} ma$$

If, however, we take 32.2 lbs. as unit of mass, that is, if we first divide m in pounds by 32.2 to get it into the new unit and then use this value in the formula, we may omit the inconvenient 1/32.2.

44. Newton's Second Law (Continued). The statement of the second law of motion is so brief that some things implied in it might easily escape notice.

1. In the statement of the law the rate of change of momentum of a body is spoken of without any reference to whether the body starts from rest or is initially in motion. Hence it is implied that *the effect of a force applied to a body is independent of the state of motion of the body when the force begins to act.* For example, gravity is a force that acts vertically downward. When a body is dropped from a height the force of gravity gives it a certain acceleration downward; if the same body be started downward with a certain velocity its acceleration downward will be the same as when the body is simply dropped, and the same will be true if the body be given an initial velocity upward or in any direction. It is found possible to play games of ball or cricket on a moving steamship; the effect of throwing the ball with a certain force or striking it with a bat is the same as when the steamship is at rest.

2. The law states how a force will affect the motion of a body, but it makes no reference to whether some other force is acting on the body at the same time or not. Hence it is implied that *each force produces its own effect independently of the simultaneous action of any other force;* and, when several forces act on a body, we may calculate the acceleration produced by each as if the other forces did not exist and then add the accelerations to find the whole effect of all the forces. This very important principle is sometimes called that of the *independence of forces.*

45. Impulse of a Force. The product of a force and the time during which it acts is called the *impulse* of the force. When a force F acts on a mass m for time t, from the formula for the Second Law of Motion, by multiplying both sides by t, we get

$$Ft = mat$$

Now at is the increase of velocity produced, and this multiplied by m is the increase of momentum. Hence *the impulse of a force equals the momentum produced by it.* If the body, starting from rest, has at time t a velocity v

$$Ft = mv$$

46. Newton's Third Law of Motion. *Action and reaction are equal and opposite.* In the statements of first and second laws of

motion forces acting on bodies are spoken of, but nothing is said as to what exerts force. This lack is supplied by the third law.

The action and reaction here referred to mean force and counter-force. The meaning of the statement is that force on any one body is exerted by some other body and this other body itself experiences an equal and opposite force exerted by the first body, the line of action of both forces being the line joining the two bodies.

In many cases the truth of this law will be recognized as being evident. For example when one presses his two hands against each other it will be admitted that the hands, if at rest, press equally in opposite directions. If one hand be pressed against a wall the same must still hold, since the wall merely takes the place of the other hand in the first illustration. But the case is not so clear when a hand is pressed against an obstacle that moves. How, it is sometimes asked, can there be motion produced if the forces are equal and opposite? The answer is that *the two forces spoken of do not act on one body;* there is one force exerted by the hand on the obstacle and the obstacle yields unless restrained by some other force; the reaction is the back pressure of the body *on the hand,* not a force acting on the body.

Consider also the forces that come into play when a horse of mass m_1 pulling on a horizontal rope of mass m_2 draws a block of mass m_3. Here there are four pairs of actions and reactions. In the first place the horse pushes against the ground and the reaction of the ground is an equal and opposite push. Let the magnitude of this horizontal action and reaction be F_1. Secondly the horse exerts a forward pull, of magnitude say F_2, on the rope and the reaction of the rope is equal and opposite. The rope exerts a force on the block and the block exerts an equal and opposite reaction, the magnitude of each being F_3. Finally there is the action and reaction between the block and the ground; let the horizontal component of this have a magnitude F_4. If there is an acceleration a, as there must be to begin the motion, F_1 is greater than F_2 by m_1a, F_2 is greater than F_3 by m_2a and F_3 is greater than F_4 by m_3a. Thus F_1 exceeds F_4 by $(m_1 + m_2 + m_3)a$ and this is, therefore, the total backward push on the ground. When the motion has become constant $a = 0$ and all the forces mentioned are of equal magnitude.

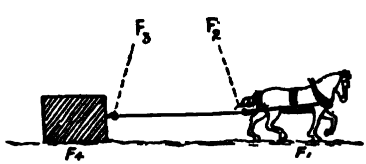


FIG. 16. Four pairs of actions and reactions.

Since a force is always accompanied by a counterforce, the two are parts or different aspects of one inseparable whole and the two together constitute what is called a *stress*. Thus every force is the partial aspect of some stress, just as a purchase and a sale are partial aspects of an exchange.

47. Force Required for Motion in a Circle. When a particle revolves in a circle it has an acceleration toward the center equal to v^2/r (§ 32) where v is the magnitude of the velocity (*i. e.*, the speed) and r is the radius. To cause this acceleration there must be a force directed toward the center, and, according to Newton's Second Law, this *centripetal* force F must be such that

$$F = m\frac{v^2}{r}$$

Against this force the particle will exert an equal and opposite reaction on the body that exerts the force toward the center. If, for example, the particle be attached by a string to the finger, the reaction will be a force acting on the finger and will be in a direction outward along the radius. This reaction is called a *centrifugal force*. Thus the centrifugal force is *not a force acting on the moving body but a reaction exerted by that body on the other body that exerts the force toward the center.* (That the above formula also applies to the motion of a *body* is shown in § 101.)

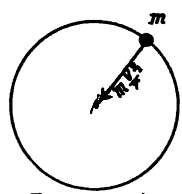


FIG. 17. A particle moving in a circle is acted on by a force toward the center.

Illustrations of the above are very numerous and a few may be mentioned. Drops of water are thrown off tangentially from a rapidly moving bicycle or carriage wheel, owing to the fact that there is not a sufficient force toward the center acting on them, and they therefore move off on a tangent in accordance with the first law of motion. A train rounding a curve presses outward on the rails and the resultant of this force and the vertical weight of the train is a force inclined to the vertical. Since it is desirable that the whole force should be perpendicular to the sleepers, the outer rail is raised. In the *Centrifugal drier* used in laundries and sugar refineries the material to be dried is placed in a perforated cylinder rotating about its axis which is vertical; the drops of water, not being held by a force directed to the center, escape through the perforations. In the

Centrifugal cream-separator, **which is a rotating vertical cylinder, both the milk and the cream tend to move as far from the axis as possible; but the milk, being the denser, exerts the more powerful tendency and therefore occupies the parts of the vessel farthest from the axis. The *flattening of the earth* at its poles is due to its axial rotation; if at rest it would be spherical; but, being in rotation, it bulges at the equator to such an extent that the restoring force of gravitational attraction equals the requisite force toward the center. The higher the speed of *belting* the less it presses on a pulley and the more liable, therefore, it is to slip; for more of the tension of the belting is called on to supply the requisite force toward the center. Watt's *governor for a steam-engine* consists of a pair of balls whirled around a vertical spindle at a rate proportional to the speed of the engine; when this speed exceeds the desired limit the outward movement of the balls acts on a steam-valve so as to decrease the speed of the engine.**

Resultant of Forces. Equilibrium.

48. Composition of Forces. Two or more forces may act on a body at the same time. For example, a body falling because of the attraction of the earth may be drawn horizontally by a stretched spring or blown by wind pressure. In such cases each force produces an acceleration independently of the action of the other forces (§ 44), and the body travels in some path with a definite acceleration, which is the resultant of the accelerations produced by the separate forces.

The resultant of two or more forces is defined as the single force which will produce the resultant acceleration. The resultant of any number of forces which act on a particle can be found by vector addition, that is by a triangle, parallelogram, or polygon construction. For a force has a certain magnitude and a certain direction and is, therefore, a vector quantity. Hence any number of forces acting on a particle may be represented by lines drawn from a point. Let $\overline{AB}$ and $\overline{AC}$ represent two forces F_1 and F_2, acting on a particle. Complete the parallelogram $ABDC$. By the Second Law of Motion the accelerations produced by F_1 and F_2 are in the directions of and proportional to $\overline{AB}$ and $\overline{AC}$ and the resultant acceleration must, therefore, be represented by $\overline{AD}$; and, since the resultant force is the force that will produce the re-

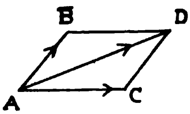


FIG. 18.

sultant acceleration, it must be in the direction $\overline{AD}$. If now we denote the acceleration produced by F_1 and F_2 by a_1 and a_2 respectively and if the resultant force and acceleration be denoted by F and a respectively, by the Second Law of Motion

$$\begin{aligned} F : F_1 : F_2 &:: ma : ma_1 : ma_2 \\ &:: a : a_1 : a_2 \\ &:: AD : AB : AC \end{aligned}$$

Hence $\overline{AD}$ represents the resultant force on the scale on which $\overline{AB}$ and $\overline{AC}$ represent the separate forces. This very important result, called the **Parallelogram of Forces**, is usually stated as follows:

If two forces acting on a particle be represented by two lines drawn from a point and if a parallelogram be drawn with these two lines as sides, the resultant will be represented by the diagonal that passes through the point.

Since, then, we may add two forces by the parallelogram method or by the triangle method (which is essentially the same), we may in the same way add a third to the resultant of these two and so on. Hence the polygon method of addition applies to forces acting on a particle.

Let θ be the angle between the directions of the force F_1 and F_2. Then as in the cases of velocities and accelerations

$$F^2 = F_1^2 + F_2^2 + 2F_1F_2 \cos \theta$$

49. Resolution of a Force into Components. Since two or more forces acting on a particle can be replaced by a single force called their resultant, a single force can be replaced by two or more forces which added geometrically give the single force. This is called the *resolution* of a *force* into *components*.

The most important case practically is when the components are at right angles to each other. When a single force is resolved into two components, the two components and the force resolved must be in the same plane. When the two components are at right angles the component that makes an angle a with the whole original force F has a magnitude $F \cos a$ and the other

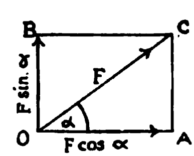


FIG. 19.

component is F sin a. The agreement as regards the signs of angles noted in § 25 applies to the present case.

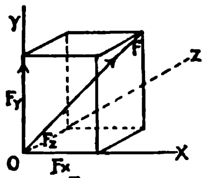


FIG. 20.

A force F may also be resolved into three components in three directions at right angles to each other. All that is necessary is to construct a right-angled parallelopiped with the line representing the force F as diagonal and with edges in the three rectangular directions. If the three directions be taken as axes of x, y and z and if the components be denoted by F_x, F_y, F_z respectively, we shall have

$$F^2 = F_x^2 + F_y^2 + F_z^2$$

50. Illustrations of the Resolution of a Force into Components. 1. The force of gravity on a body of mass m acts vertically downwards and in absolute units equals mg. If a body is not free to move vertically but is free to move in some other direction, the only part of gravity that can affect the motion is the component in that direction. For instance if a body be placed on a smooth plane inclined at an angle i to the horizontal the force of gravity, mg, may be resolved into a component $mg \sin i$ down the plane and a component $mg \cos i$ perpendicular to the plane. The latter component will produce pressure on the plane but will not affect the motion down the plane, which will depend only on the former component mg sin i. If the plane be not perfectly smooth there will also be a force of friction, say F, parallel to the plane and the resultant force down the plane will be $(mg \sin i - F)$.

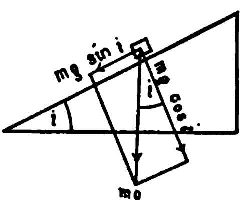


FIG. 21.

2. A sail-boat effects a double resolution of the wind pressure. The component of the wind pressure W parallel to the plane of the sails has very little effect; the component, say F, perpendicular to the sail is the effective component. Again F may be resolved into a component perpendicular to the keel and a component f parallel to the keel. The former produces a small sidewise motion or lee-way while the latter, being in the direction in which the boat is most free to move, is the effective component.

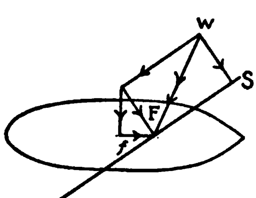


FIG. 22.

3. In the case of a kite the component of the wind pressure parallel to the surface of the kite has no effect; the component perpendicular to the kite when the kite has risen to the proper level is equal and opposite to the resultant of the pull T of the cord and the weight mg of the kite.

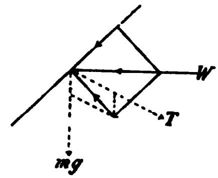


Fig. 23.

51. Analytical Method of Compounding Forces. A simple and general method of finding the resultant of a number of forces in a plane is to resolve each in two directions at right angles and then add all these components. Thus let $F_1, F_2 \cdots$ be the forces, acting on a particle at O. Take any two convenient rectangular directions Ox and Oy and let the angles $F_1, F_2 \cdots$ make with Ox be $\alpha_1, \alpha_2 \cdots$ respectively. Then F_1 is equivalent to $F_1 \cos \alpha_1$ along Ox and $F_1 \sin \alpha_1$ along Oy and so for the other forces. Let the sum of the components along Ox be denoted by X and the sum of the components along Oy by Y. Then

$$X = F_1 \cos \alpha_1 + F_2 \cos \alpha_2 + \cdots = \Sigma F \cos \alpha$$
$$Y = F_1 \sin \alpha_1 + F_2 \sin \alpha_2 + \cdots = \Sigma F \sin \alpha$$

We have thus replaced the forces $F_1, F_2 \cdots$ by X along Ox and Y along Oy. The resultant of X and Y is the resultant of F_1, F_2, etc. Let the magnitude of the resultant be R and let it make an angle θ with OX. Then

$$R^2 = X^2 + Y^2$$
$$\tan \theta = \frac{Y}{X}$$

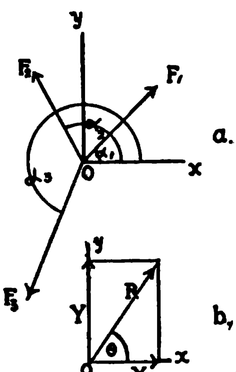


Fig. 24. Analytical method of compounding forces.

These formulæ give the magnitude and the direction of the resultant. In using this method it must be remembered that, when we substitute for each angle α its numerical value, we must call the angle positive if it is measured in the direction regarded as positive, say the counter-clockwise direction; if measured in the opposite direction it must be regarded as negative.

The angle θ that the resultant makes with Ox is found from

its tangent. When the tangent is positive it shows that the angle is between 0 and 90° or between 180° and 270°. To decide between these two, note that the signs of the values of X and Y must be either both positive or both negative, since the tangent is positive. If both are positive θ is between 0° and 90°; if both are negative it lies between 180° and 270°. The reader should have no difficulty in completing the reasoning for the case in which $\tan \theta$ is negative.

When X and Y are both zero, that is, when the sum of the components in each of two directions at right angles is zero, R is also zero. Conversely, when R is zero X and Y must also each be zero, since the square of a number cannot be negative.

When the forces to be compounded are not all in one plane we may take three directions, Ox, Oy, Oz, at right angles and resolve each force into components in these three directions. Denote the sum of the components along Ox by X, that along Oy by Y, and that along Oz by Z and let the resultant be R. Then

$$R^2 = X^2 + Y^2 + Z^2$$

If $X = 0$, $Y = 0$ and $Z = 0$, then $R = 0$. The converse is also true, since X^2, Y^2 and Z^2 must be either positive or zero.

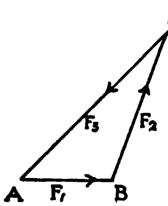


FIG. 25.

52. Equilibrium of Forces Acting on a Particle. *When the resultant of the forces acting on a particle is zero the forces are said to be in equilibrium,* that is, in a state of balance so that they do not change the motion of the particle.

When two equal and opposite forces act on a particle they are in equilibrium, for their resultant is zero. Conversely, if two forces are in equilibrium they must be equal and opposite, for otherwise their resultant could not be zero.

When *three forces acting on a particle are in the direction of and proportional to the sides of a triangle taken in order, they are in equilibrium.* For if the three forces F_1, F_2, F_3 be in the direction of and proportional to $\overline{AB}$, $\overline{BC}$, $\overline{CA}$, the resultant of F_1 and F_2 will be represented by $\overline{AC}$ and the resultant of forces represented by $\overline{AC}$ and $\overline{CA}$ is zero. The converse of this proposition is that, *if three forces acting on a particle be in equilibrium, and if any*

triangle be drawn with its sides respectively in the directions of the forces, the forces will be proportional to the sides of this triangle. To prove this statement let us suppose that $\overline{AB}$ and $\overline{BC}$ are any two lines in the direction of and proportional to two of the forces F_1 and F_2. Then a force represented by $\overline{AC}$ is equivalent to F_1 and F_2 taken together. Hence since the forces are in equilibrium the third force F_3 must be in the direction of and proportional to $\overline{CA}$. Now *any other triangle* such as $A'B'C'$ with sides in the directions of F_1, F_2, F_3 respectively, that is, in the direc-

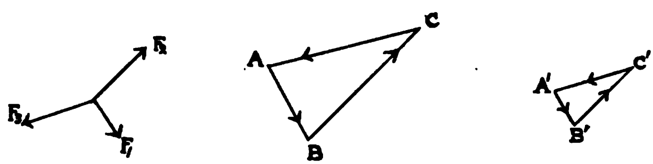


FIG. 26.

tions of $\overline{AB}$, $\overline{BC}$, $\overline{CA}$, respectively, must be similar to ABC. Hence its sides must be proportional to the sides of ABC, that is, to F_1, F_2, F_3 respectively. This converse proposition is very important, for, when we know the directions of three forces that are in equilibrium, we can find the relative magnitudes of the forces by constructing a triangle with its sides in the directions of the forces.

When ***any number of forces*** acting on a particle are in the directions of and proportional to the sides of a closed polygon taken in order, they are in equilibrium; for the resultant is zero. The converse of this proposition, for more than three forces, is not true; for polygons are not necessarily similar when their respective sides are parallel.

When ***any number of forces*** are such that the sum of their components in each of three directions at right angles is zero, they are in equilibrium. This is evident from § 51, for when X, Y and Z are all zero R must also be zero. Conversely, when any number of forces are in equilibrium, the sum of their components in ***any*** direction equals zero; for we may take this direction as one of three at right angles; and, since R is zero, the sum of the components in each of these three directions is zero (§ 51).

Work and Energy.

53. Work. When a force acts on a body *the product of the force by the distance through which it acts in the direction of the force* is, as we shall see later, a very important quantity and is called the *work* performed by the force. Thus when a force applied to a heavy body raises it a certain vertical distance, work is performed by the force, the amount of the work being the product of the force and the distance of ascent; and when a horizontal force draws a body horizontally the work is the product of the force and the horizontal distance.

The phrase "in the direction of the force" that occurs in the definition of work should be carefully noted. When there is no motion in the direction of a force no work is performed by that force. For instance, a travelling crane may by its chains exert an enormous force in sustaining a heavy body and it may move the body through a great distance horizontally, but the force exerted by the chains will do no work if there is no vertical motion. If a force F acts constantly on a body while the body moves a distance AB which is not in the direction of the force, to get the work performed we must take the projection of AB on the line of action of the force and multiply the projection by the force. If θ is the angle between AB and the direction of the force, the projection, AC, of AB on the line of action of the force

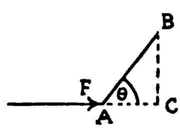


FIG. 27.

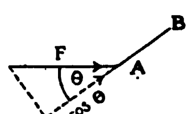


FIG. 28.

is $AB \cos \theta$ and the work performed is $F \cdot AB \cos \theta$. This at once suggests another method of calculating the work performed, for $F \cdot AB \cos \theta$ is the same as $F \cos \theta \cdot AB$ and $F \cos \theta$ is the component of F in the direction of AB. Thus the work performed by a force is also the product of the total distance by the component of the force in the direction of motion.

Work, or the product of force and distance, must be carefully distinguished from the impulse of a force, which is the product of a force and the time during which it acts (§ 45). Given a force

and the distance through which it acts, we do not need to know the time in order to calculate the work.

54. Positive and Negative Work. Forces always exist in pairs of equal and opposite forces (§ 46). Hence, when a force applied to a body does work by moving the body in the direction of the force, it must at the same time overcome an opposing force or reaction. The applied force in this case does positive work, since the motion is in the direction of the applied force. This work is done against the reaction, or we may say that the reaction does negative work, since the motion is in the opposite direction to the reaction.

The nature of the reaction is different in different cases. A horse attached to a wagon is doing work against the force of friction when the wagon is moving uniformly, or we may say that the force of friction is doing negative work. In starting the wagon into motion the horse does work against the inertia of the wagon and also of the horse, in addition to the work it does against friction. When a body is moving in one direction and a force is suddenly applied in the opposite direction the body does positive work against the force, which in this case does negative work.

55. Units of Work. The unit of work is *the work done by the unit force in acting through unit distance.* When the dyne is taken as unit of force and the cm. as unit of length, the unit of work is that performed by a dyne acting through a cm. and is called an **erg.** Since this is a very small unit, a multiple of it, namely 10,000,000 ergs, is frequently used and is called a **joule.**

When the weight of a pound is taken as unit of force and the foot as unit of length, the unit of work is the work done by a force equal to the weight of one pound when it acts through one foot and is called a *foot-pound.*

56. Diagram of Work. When a force is constant, to find the work it does we multiply the magnitude of the force by that of the displacement; but, when a force is variable, some other method has to be adopted. One way is to divide the whole displacement up into small parts and multiply each small part by the force at the middle of the small displacement and then add all the products. By taking the parts small enough we may get the work as accurately as may be desired. A graphical method is often preferable.

It is entirely similar to the method used in finding the distance a point travels when it has a variable velocity (§ 22). Let OA be a line that represents on some scale the whole displacement measured in the direction of the force. Divide OA into a very large number of very small equal parts. At O erect a perpendicular OB to represent on some scale the force at the beginning of the first part; erect similar perpendiculars to represent the magnitude of the force at the beginning of the other parts and through the ends of these perpendiculars draw a smooth curve BC. If we calculated the work done in a small displacement ab by taking for the force its value at the beginning of ab, the result would be too small; and if we made the calculation by taking the value of the force at the end of ab the result would be too large, and similarly for all the other intervals. By continuing the reasoning as in § 22 we find that the actual work done is represented by the area $OBCA$.

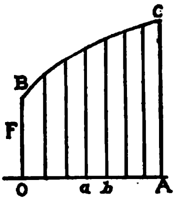


FIG. 29. Diagram of work.

Thus to find the whole work we need only to measure the area of the figure and then allow for the scale on which it is drawn. If each unit of length along OA stands for m units of length in the displacement and if each unit of length along OB stands for n units of force, each unit of area will stand for mn units of work, and the whole area multiplied by mn will give the whole work.

When the curve of force is a straight line the area may be readily calculated. For example let us calculate the work done in stretching a spring. In this case it is known that the force that is needed to keep a spring stretched is proportional to the amount of the stretch or increase of length (provided this be not so great as to permanently lengthen the spring). Hence if the spring is stretched by an amount x the force applied to it is kx where k is a constant. If, then, a curve be drawn with the values of kx as ordinates and the values of x as abscissæ, this curve will be a straight line (Fig. 30), which will pass through the origin, since kx is zero when x is zero. To find the work done in increasing the amount of the stretch from x_1 to x_2 where $OL = x_1$ and $ON = x_2$, we must find the area $PLNQ$. Now this is equal to $\frac{1}{2}LN(PL + QN)$. Hence the work done is $\frac{1}{2}(x_2 - x_1)(kx_2 + kx_1)$ or $(\frac{1}{2}kx_2^2 - \frac{1}{2}kx_1^2)$. This is also the work the spring will do in

contracting, since at each step the force of contraction is equal to the force required to stretch. If the initial stretch be zero, $x_1 = 0$, and the work required to stretch by the amount x_2 is $\frac{1}{2}kx_2^2$. While we have referred especially to the force exerted by a spiral spring, the above proof and formula evidently apply to the work done by any force that is proportional to displacement. These we shall find later are very numerous.

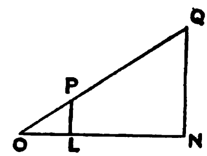


FIG. 30.

57. Power or Activity. The *rate* at which an agent works or the number of units of work performed per unit time is called the *power* or activity of the agent. In C.G.S. units the unit of activity is that of an agent that does one erg per second. As this unit is extremely small, the unit employed for most scientific purposes is 10^7 ergs per second or one joule per second and is called the *watt;* a still larger unit is the *kilowatt* which equals one thousand watts.

The unit largely employed for engineering purposes is the *horse-power,* which is the power of an agent that does 550 ft. pounds per second or 33,000 ft. pounds per minute.

58. Kinetic Energy. Consider the case of a constant force F acting on a body of mass m which is not acted on by any other force. F will cause an acceleration a such that $a = F/m$. If the velocity increases from u to v in a distance s

$$v^2 = u^2 + 2as$$

To get the work performed by F we substitute the value of a and thus get

$$Fs = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mu^2$$

One half the product of the mass of a body by the square of its velocity is called the kinetic energy of the body. We may therefore state the above expression thus:

Work done on body = gain of kinetic energy;

but it must be remembered this is only for the case in which the force acts on a body which is otherwise free.

We also supposed that the force is a constant one; but we may readily extend the conclusion to the case in which the force is not constant. For we may divide the whole distance into very small parts in each of which the force is practically constant and in each the work done will be equal to the gain of kinetic energy. If we denote the small movements by s_1, s_2, etc., and the forces in these by F_1, F_2, etc., respectively, the total work done will be the sum $(F_1s_1 + F_2s_2 + \cdots$ etc.) which we may abbreviate to ΣFs. The total gain of kinetic energy will be the final kinetic energy minus the initial kinetic energy. Hence if the initial velocity is u and if the final velocity is v

$$\Sigma Fs = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mu^2$$

If the body starts from rest so that $u = 0$ the work done while the body acquires the velocity v will be $\frac{1}{2}mv^2$.

We may also reverse the circumstances and enquire what work a body in motion can do if it meets an opposing force and is brought to rest. Suppose that it exerts a constant force F. Then the opposing force, that is the force applied to the body, will be $-F$ and the acceleration will be a negative one equal to $-(F/m)$, and since as before

$$v^2 = u^2 + 2as$$

we get by substitution

$$Fs = \tfrac{1}{2}mu^2 - \tfrac{1}{2}mv^2$$

Thus the work done *by the body* against the resistance is equal to the *loss* of kinetic energy of the body. If the force overcome is not constant, as when the body compresses a spring, we may as before divide the whole distance into small parts and summate. Thus if the initial velocity be u and the final velocity v we shall have the expression ΣFs for the whole work. Hence

$$\Sigma Fs = \tfrac{1}{2}mu^2 - \tfrac{1}{2}mv^2$$

If the motion continue until the body is brought to rest, v will then be zero and we shall have the result that *the initial kinetic energy of the body is the work it can do before it is brought to rest.* It should be noticed that in the above u and v stand for the

magnitudes of the respective velocities, *i. e.,* the speeds (§ 16). The kinetic energy of a body depends on the square of the magnitude of its velocity and is the same no matter what the direction of motion, that is, kinetic energy is a scalar quantity; to the kinetic energy of one body we may add the kinetic energy of another body and the sum will be the total kinetic energy of both bodies.

Since a force does no work when it is always at right angles to the direction of motion, it follows that when a body is acted on by a single force at right angles to the direction of motion the kinetic energy of the body remains constant. Thus when a body rotates in a circle under the action of a single force directed toward the center the force does no work and the kinetic energy of the body is constant.

Kinetic energy and work are equivalent quantities; hence the units of kinetic energy are the same as the units of work.

59. Kinetic Energy and Gravity. The force of gravity on a body is for small distances above the surface of the earth a constant force. If a body at a height H above the earth's surface has a velocity v vertically downwards, when it has fallen so that its distance above the surface is h gravity will have done an amount of work $mg(H-h)$; and, if the velocity of the body be then V, its kinetic energy will have increased from $\frac{1}{2}mv^2$ to $\frac{1}{2}mV^2$. Hence

$$mg(H-h)=\tfrac{1}{2}mV^2-\tfrac{1}{2}mv^2$$

If on the other hand the body be projected upward with a velocity V from a height h, it will be opposed by the force mg and the work it will do against gravity in rising to a height H will be $mg(H-h)$. If its velocity at the height H be v, its loss of kinetic energy will be $(\frac{1}{2}mV^2-\frac{1}{2}mv^2)$. Equating the work done against gravity to the loss of kinetic energy we get the same equation as above.

In the preceding we have supposed the motion to be vertical; but the result will be unchanged if the motion is not vertical, provided no force except gravity act on the body in the direction of its motion. Any force perpendicular to the motion will do no work and cause no change of kinetic energy. Suppose, for example, the body slides down a smooth plane through a distance s along the plane. Now we have already shown that it acquires the

same velocity as if it fell vertically a distance equal to the height of the plane (§ 35). Then if H be the height of the top of the plane and h that of the bottom the general equation above will still hold. The same is true if the descent is along a smooth curve; for a curve may be regarded as made up of very short straight parts to each of which the principle stated will apply. These results are now readily understood by considering the work performed by gravity. For the total amount of motion in the direction of the whole force of gravity is $(H-h)$. Thus the gain of kinetic energy in the descent from the higher level to the lower must be the same as if the fall were vertical.

60. Kinetic Energy and Elasticity. When a body is acted on by the force due to a stretched spiral spring, the spring will do work on the body if the spring is contracting and the body will do work against the force of the spring if it is moving so as to further stretch the spring. Let us first suppose that the body is moving toward the spring with a velocity v, the spring being at that moment stretched to an amount X beyond its normal or unstretched length. While the spring is contracting the velocity of the body will be constantly increasing. Let the velocity be V when the spring has contracted so that its stretch is decreased to x. In this time (§ 56) the spring will have done an amount of work $(\frac{1}{2}kX^2-\frac{1}{2}kx^2)$ and, since this must equal the increase of kinetic energy of the body,

$$\tfrac{1}{2}kX^2-\tfrac{1}{2}kx^2=\tfrac{1}{2}mV^2-\tfrac{1}{2}mv^2$$

We may also suppose the case reversed, that is, we may suppose the body to be moving away from the spring with a velocity V when the stretch of the spring is x. Then the velocity of the body will decrease; and, if it be v when the stretch of the spring is X, work $(\frac{1}{2}kX^2-\frac{1}{2}kx^2)$ will have been done against the spring and the decrease of the kinetic energy of the body will be $(\frac{1}{2}mV^2-\frac{1}{2}mv^2)$. Equating these we get the same equation as before.

61. Potential Energy. We shall now consider the two illustrations just given from another point of view. In the case of a body projected vertically upward there is a loss of kinetic energy equal to mg multiplied by the height of ascent; and, if the body be allowed to descend again, the same amount of work will be

performed by gravity and the body will regain its lost kinetic energy. Thus at the higher level the body (or rather the body and the earth regarded as one system) has an advantage of position that is equivalent to a certain amount of kinetic energy lost and this advantage of position is measured by $mg(H-h)$. This, since it is equivalent to a certain amount of kinetic energy, is called *potential energy.* Thus it follows that *the sum of the kinetic energy and the potential energy is a constant,* a fact brought out more clearly by writing the equation of § 59 thus:

$$\tfrac{1}{2}mV^2 + mgh = \tfrac{1}{2}mv^2 + mgH$$

Here mgh is the increase of the potential energy when the body is raised from the arbitrary zero level (*e. g.,* sea-level) from which h is measured to the height h, and a similar statement applies to mgH. When the body is at the zero level, it and the earth still possess potential energy, since work could be obtained by allowing the body to fall down a vertical shaft.

Again in the case of the work done against a spring by a moving body there is a decrease of kinetic energy, and this decrease is equal to the work done against the spring. If the motion be reversed, the lost kinetic energy will be regained. Thus when the stretch of the spring increases from x to X it acquires a capacity for doing work of the amount $(\tfrac{1}{2}kX^2 - \tfrac{1}{2}kx^2)$, equal to the kinetic energy lost by the body; and the spring yields up this capacity for doing work in restoring the kinetic energy of the body. Writing the equation of § 60 in the form

$$\tfrac{1}{2}mV^2 + \tfrac{1}{2}kx^2 = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kX^2$$

we see that the sum of the kinetic energy of the body and the work the spring can do in contracting to its unstretched length is a constant. The work the spring can do in contracting to its unstretched length is the potential energy of the spring.

In the case of a body separated from the earth the potential energy of the body and the earth depends on their *relative position,* and in the case of the energy of the spring the potential energy depends on the relative positions of the parts of the spring. Hence we may say that *potential energy is the capacity a body or*

system of bodies has for doing work in virtue of the relative positions of its parts.

In the case of potential energy we cannot give any universal formula by which it can be calculated as we can in the case of kinetic energy. In each case of potential energy we must calculate how much work the body or system can do in passing from one state to another and take this as the difference of the potential energy of the body or system in the two states. For any one particular case of potential energy we may deduce a special expression for its amount, such as those given above for gravity and elasticity.

62. Interchanges of Kinetic and Potential Energy. We have considered somewhat fully two cases of the interchange of kinetic and potential energy, namely, those of gravity and elasticity, because these are typical and are easily worked out by elementary methods. Such interchanges are common in nature and in industry and a few may be briefly stated.

(*a*) *Change from Kinetic to Potential.* When a block of wood is split by a wedge or axe, the axe or sledge hammer loses kinetic energy and potential energy of separation of the particles of wood is produced.

As the distance of the earth from the sun increases from midwinter to midsummer the speed of motion and the kinetic energy decrease and the potential energy of separation increases.

(*b*) *Change from Potential to Kinetic.* A clock-weight or watch-spring when wound up has potential energy, and this changes to kinetic energy of the pendulum or balance wheel, which would otherwise come to rest.

A bent bow has potential energy due to the change of position of the particles of the bow and the forces between them. As it unbends it loses this potential energy and the arrow gains kinetic energy.

Water in a lake or above a dam has potential energy; when allowed to escape to a lower level it loses part of its potential energy and either gains kinetic energy itself or, if it acts on a water-wheel or turbine, it imparts kinetic energy to the latter.

(*c*) *Periodic Interchanges.* In any case of vibration *energy continually changes from the kinetic to the potential form and*

back again. Thus in the vibration of a pendulum, at the bottom of the arc of vibration the potential energy is reduced to zero, and the kinetic energy is at a maximum, while at the end of the arc of vibration the kinetic energy is zero and the potential energy has increased to a maximum. Similar statements apply to the vibration of a tuning fork, a violin string, a body attached to the end of a wire and vibrating torsionally, the oscillations of the balance-wheel of a watch, and so on.

63. Two Kinds of Forces. In the preceding we have seen that, when the forces acting between bodies are forces of gravity or forces of elasticity, their action leaves the total kinetic and potential energy of the bodies unchanged, or, as it is usually stated, when only such forces act the total kinetic and potential energy is *conserved.* Forces whose action between bodies does not cause a change of the total kinetic and potential energy of the bodies are called *conservative forces,* and any system of bodies between which the forces are wholly conservative is called a *conservative system.*

In contrast with these conservative forces stands such a force as friction. A moving body opposed by friction loses kinetic energy as its velocity decreases but it does not at the same time gain potential energy to an equivalent extent. Thus a body started up a rough inclined plane with a certain velocity will not reach as high a level as it would reach if the plane were smooth, and it will not have as much potential energy when it reaches its highest point. Moreover, its descent will be further opposed by friction and its store of kinetic and potential energy will thereby be further reduced. Friction, then, is a non-conservative force since, when in action, it causes a permanent decrease of the kinetic and potential energy of a system.

The reason why such a force as gravity has no effect on the sum total of kinetic and potential energy is easily seen. At a certain distance of a body from the earth the force between the two depends only on their distance apart, and is independent of the way in which they are moving. Hence, when they are moving away from each other and are a certain distance apart, they are losing kinetic energy at a rate exactly equal to the rate at which they regain kinetic energy when, at the same distance of separa-

tion, they are moving toward one another. Thus forces of gravity between bodies *depend only on the relative positions of the bodies.* The same is true of the forces between the parts of an elastic spring and this accounts for the fact that such forces of elasticity are also conservative; in fact it is the fundamental characteristic of all conservative forces. But a non-conservative force, such as friction, depends on the way in which a body or a system of bodies is moving; it is always opposed to the direction of relative motion of bodies in contact; hence it causes a diminution of the kinetic energy of the bodies in whichever direction motion is taking place.

X **64. The Conservation of Energy.** We have seen in the preceding that *under certain conditions the total kinetic and potential energy of a system is constant or is conserved.* The conditions referred to are two, (1) the system must not receive energy from or give energy to any outside bodies, (2) the forces between the parts of the system must be wholly conservative. In reality no system wholly satisfies these conditions. No system is wholly *isolated* in the sense implied in the first condition; and non-conservative forces, such as friction, are never quite absent. But in many cases these conditions are very nearly satisfied. The solar system, consisting of the sun, planets, and moons, is practically isolated; and, while there are internal frictional forces such as those of the tides, the work they do is so small compared with the total energy of the system that their effects in reducing the kinetic energy of the whole have not yet been detected with certainty. Again the system consisting of the earth and a body vibrating as a pendulum in a vacuum is practically an isolated system free from frictional forces and the total kinetic and potential energy is very nearly constant; the same is true of a heavy body attached to a spring and vibrating in a vacuum. When, as in cases like these, the conditions are sufficiently nearly satisfied the principle of the constancy of kinetic and potential energy will often lead to valuable results.

In an isolated system in which there are non-conservative forces such as friction, energy is expended in doing work against these forces; and, if to the sum of the kinetic and potential energy we add the work done against non-conservative forces, the sum will be constant. But what becomes of the energy so expended? For

long it was supposed to be wholly lost. It was of course known that heat was produced when work was done against friction; but heat was supposed to be a form of matter. But about 1840 the view was advanced that heat, instead of being a form of matter, is a form of energy as this word is now defined, and this led to the discovery of the **Law of Conservation of Energy**, which is treated fully later under "Heat."

KINEMATICS.

ROTATION.

65. Angular Displacements. In §§ 9–35, we studied the motion of translation of a point as a preliminary to the study of the effect of forces on the motion of particles and of bodies moving without rotation. We shall now consider the motion of bodies in rotation as a preliminary to studying the effects of forces on the motion of rotation of bodies.

The motion of a body is one of rotation when each point in the body moves in a circle the center of which is on a straight line called the axis of rotation. All points in the body turn in any time through equal angles and the angle described in any time is called the *angular displacement* of the body in that time. Its magnitude may be stated either in degrees or in radians, but the latter method is in many ways the more convenient for the present purposes.

A radian is an angle which is subtended at the center of a circle of radius r by an arc of length r. An angle subtended in a circle of radius r by an arc of length a contains a/r radians. Hence 4 right angles equal 2π radians and one radian equals $360/2\pi$ degrees or $57°.29578$.

66. Angular Velocity. The rate of rotation of a body is called its *angular velocity*. When the angular displacements of a body in all equal times are equal, the velocity is a constant angular velocity, and the magnitude of the angular velocity is *the angle through which the body turns in unit time*. If the angle is reckoned in radians and the second is taken as unit of time, the magnitude of the angular velocity is the number of radians described in one second. The unit of angular velocity is one radian per second.

If the velocity is not constant, as, for example, when a fly-wheel is being set in motion or stopped, the angular velocity or rate of angular displacement is defined in the same way as in the analogous case of variable linear velocity (§ 19), that is to say, we must take the average angular velocity in a short time and then suppose this time indefinitely decreased, so that the average angular velocity approaches a limiting value, which is the *instantaneous angular velocity.*

67. Angular Acceleration. The rate of increase of the angular velocity of a body is called its *angular acceleration.* When the angular velocity increases by equal amounts in equal times, the angular acceleration is constant and its magnitude is *the increase of angular velocity in unit time.* If we denote the angular acceleration by α, the increase of angular velocity in each second is α and the increase in t seconds is αt. Hence, if at the beginning of an interval of time t the angular velocity is ω_0 and at the end of the interval it is ω,

$$\omega = \omega_0 + \alpha t \tag{1}$$

In this time the body has turned through a certain angle say ϕ. To find the magnitude of ϕ we may represent the varying values of the angular velocity by means of a curve of angular velocity as we did in the similar case of a varying linear velocity (§ 27), and the area of the diagram will represent the angle ϕ. The two diagrams would have precisely similar properties, the only difference being that in one case we would speak of linear displacement, s, linear velocity, v, linear acceleration, a, while in the other case we would speak of angular displacement, ϕ, angular velocity, ω, and angular acceleration, α. Hence, when the angular acceleration is constant, the formula for ϕ, which must be precisely similar to (2) of § 27, is

$$\phi = \omega_0 t + \tfrac{1}{2}\alpha t^2 \tag{2}$$

By elimination of t between (1) and (2) we get

$$\omega^2 = \omega_0^2 + 2\alpha\phi \tag{3}$$

68. Angular Velocity and Linear Velocity. When a point re-

volves at a constant rate in a circle, its motion may be described either by means of its angular velocity, ω, or by its linear velocity, v, along the tangent, and there is a simple relation between the two. Let the radius of the circle be r and let the time in which the point moves from P to Q be t. Denoting the length of the arc PQ by s and the angle POQ by ϕ, we have from the definitions of linear and of angular velocity

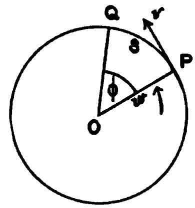


FIG. 31.

$$s = vt; \ \phi = \omega t$$

Now in radian measurement $\phi = s/r$. Substituting in this the above values of s and ϕ we get

$$\omega = v/r$$

Thus the relation between angular velocity and linear velocity when a point rotates in a circle is the same as the relation between an angle and the arc which it subtends.

The above relation is important. More briefly stated, the proof of it amounts to this: v is the length of arc described per second; hence v/r is the angle described per second in radian measurement, that is the angular velocity.

When a point describes a circle with variable speed, the above relation holds true, with the understanding that ω and v are the instantaneous values of the angular and the linear velocity respectively. The proof is the same as above, t being taken as a very short interval.

When a body rotates about an axis with angular velocity ω, a point in the body describes a circle of radius r, and r is different for points at different distances from the axis. If r and r' are the respective distances of two points from the axis and v and v' their respective linear velocities, $v = r\omega$ and $v' = r'\omega$. Hence $v:v'::r:r'$. When the axis about which a body rotates varies from moment to moment, the above relation is true of the values of ω, v, and r at any moment. For example the wheel of a moving wagon or bicycle is always in contact with the road and the point

of contact is at any moment the point about which the whole wheel is at that moment rotating. Now the top of the wheel is twice as far from the ground as the center of the hub and must, therefore, have twice as great a linear velocity.

69. Angular Acceleration and Linear Acceleration. When a point revolves in a circle with changing angular velocity it has an angular acceleration, say α. The speed of the point along the tangent increases with an acceleration, say a. The same relation holds between a and α as between v and ω (§ 68). For if ω is the angular velocity at the beginning of a short time t and v the linear speed at this time, $v=\omega r$; at the end of the time t the angular velocity is $(\omega+\alpha t)$ and the linear speed is $(v+at)$. Hence $(v+at)=r(\omega+\alpha t)$. Subtracting the former equation from the latter and cancelling t we have

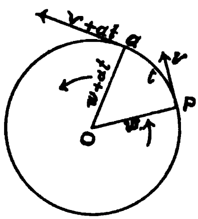


FIG. 32.

$$a=r\alpha$$

More briefly stated, a is the added linear speed per unit time and α the added angular velocity per unit time, and the relation between angular velocity and linear speed must hold true of these increases.

It should be carefully noted that a here means the rate of change of speed along the tangent. Since the direction of the velocity is also changing this cannot be the only acceleration. In fact as we have already seen (§ 32) there is in all cases of motion in a curve a linear acceleration towards the center equal to v^2/r, or, as we may now write it, $\omega^2 r$, since $v=r\omega$.

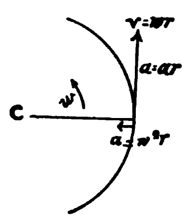


FIG. 33.

The above relations, which are very important, are summarized in figure 33.

70. Resolution of Uniform Circular Motion into Components. The velocity of a point which moves in a plane may always be resolved into two components at right angles in that plane, and the same is true of the acceleration of the point. When a point moves in a circle with constant angular velocity ω (reckoned counter-clockwise) we may resolve the linear velocity and the accelera-

tion of the point into two components along two rectangular diameters AA' and BB'. We shall, for a future purpose, consider one of these components, say that along AA'.

The linear velocity of P is ωr along the tangent and if at a certain moment the radius CP makes an angle θ with CA, the direction of this velocity makes an angle $((\pi/2)+\theta)$ with the positive direction of CA. Hence the component velocity along CA is $\omega r \cos((\pi/2)+\theta)$ or $-\omega r \sin\theta$. Let PM be perpendicular to CA and denote CM by x. Then $\sin\theta = \pm(\sqrt{r^2-x^2})/r$. Hence the component along CA of the velocity of P is $\pm\omega\sqrt{r^2-x^2}$, the sign being negative when the velocity is from A to A' and positive when it is from A' to A.

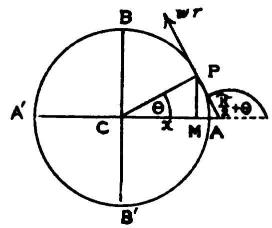


FIG. 34.

The acceleration of P is $\omega^2 r$ along PC, which makes an angle θ with the direction of AC. Hence the component acceleration along AC is $\omega^2 r\cos\theta$ or $\omega^2 x$; reckoned in the direction CA the component acceleration is therefore $-\omega^2 x$.

Since M is the projection of P, M moves backward and forward along $A'A$ as P revolves in the circle, and the component velocity and acceleration found above are the velocity and acceleration respectively of M.

71. Graphical Representation of Angular Quantities. An angular displacement is of a certain magnitude and about a certain axis. Given the axis, the direction of rotation around it and the magnitude of the angular displacement, we know everything about it. Now all these can be represented graphically by *a length marked off on the axis* so as to represent to some scale (*e. g.*, a cm. per radian) the magnitude of the angular displacement. There must also be some agreement as to which direction along the axis shall represent a certain direction of rotation about the axis. The rule usually adopted for this purpose is called the "right-handed screw rule," namely, *let the direction along the axis of the line that represents an angular displacement be related to the direction of the rota-*

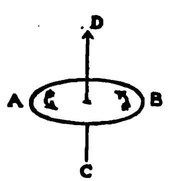


FIG. 35. A rotation indicated by the arrows is represented by *CD*.

tion as the direction of translation is to the direction of rotation of an ordinary (right-handed) screw. For example, a line to represent the angular displacement of the earth in 24 hours due to its rotation about its axis would be drawn from the center toward the N. pole. Two lines to represent the angular displacements of the hands of a watch in one hour would be drawn through the center of the face toward the back and the one for the minute hand would be twelve times as long as the one for the hour hand.

A line to represent an angular velocity would be laid off on the axis of rotation according to the above rule, and a line to represent an angular acceleration would be drawn in the same way.

A directed line that represents according to the above agreement an angular displacement is a vector, since it has both magnitude and direction; but it differs from vectors that represent linear displacements in the fact that it must in any diagram be located on a certain line, namely the line that stands for the axis of rotation. Such a vector is therefore called a *localized vector* or *rotor.* Two parallel and equal vectors, not in the same line, do not represent the same angular displacement, since the rotations they represent are about different axes.

72. Addition of Angular Velocities and Accelerations about Intersecting Axes. A body may have two or more simultaneous angular velocities. For example, suppose a bicycle wheel while rotating about its axis is mounted on a horizontal platform which is kept in rotation about a vertical axis. At any moment the wheel has two component angular velocities about intersecting axes. Each may be represented by a vector drawn from the center according to the rule stated in § 71. We may then add these two vectors by the parallelogram method and the diagonal will represent in magnitude and direction the resultant angular velocity at the moment in question. (This is fully proven in advanced treatises.)

Since angular accelerations are increments of angular velocities per unit time, we may add them as we add angular velocities.

It follows from the above that angular velocities about *intersecting axes* may be compounded and resolved by the methods applicable to linear velocities (§§ 23–25) and a similar statement holds true for angular accelerations about intersecting axes.

DYNAMICS.

CENTER OF MASS.

73. General Description of Center of Mass. When the motion of a rigid body is one of translation without rotation, all points in the body move exactly in the same way, and in describing or calculating the motion any point in the body may be taken as representing the whole body. When the motion is one of translation combined with rotation, different points in the body move differently and there is no one point the motion of which completely represents the motion of the whole body. There is, however, in any body one particular point which, for many purposes, may be taken as representing the body, so that for these purposes the body may be regarded as concentrated to a particle at that point. This point, which we shall define presently, is the *center of mass* of the body. For instance, let a uniform circular disk be tossed into the air; it will be seen that the center of the disk moves like a particle either in a straight line or in a parabola, while other points in the disk rotate around it. If the disk is loaded with lead on one side it will be some other point, not the geometrical center, that will show this property.

If a body wholly free were struck a blow at random, it would start with both translation and rotation; but if the blow were applied at the center of mass or in a line through the center of mass, the motion would be one of translation without rotation.

The center of mass is thus seen to be a point of great importance in describing or calculating the whole motion of a body. In what follows we shall define the center of mass and show how its position may be calculated. Then from the definition we shall deduce the above and other properties.

74. Definition of Center of Mass. 1. *Of Two Particles.* Let the particles be m_1 at P_1 and m_2 at P_2. Let C_1 be a point that divides P_1P_2 inversely as the masses of the particles, that is, such that

$$\frac{C_1P_1}{C_1P_2} = \frac{m_2}{m_1}$$

C_1 is the center of mass of m_1 and m_2.

2. *Of Three Particles.* Let the particles be m_1 and m_2 as above

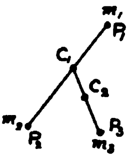


FIG. 36.

and m_3 at P_3. Suppose m_1 and m_2 replaced by (m_1+m_2) at C_1 and let C_2 be a point in C_1P_3 such that

$$\frac{C_2P_3}{C_2C_1} = \frac{m_1+m_2}{m_3}$$

C_2 is the center of mass of m_1, m_2 and m_3.

3. *Of Any Number of Particles or of a Body.* Proceeding as above we get the center of mass C of any number of particles, and the same will apply to a body of any form, since it may be divided up into a large number of small parts.

We shall show in the next section that the point to which such a process leads is independent of the order in which the particles are taken.

75. Distance of the Center of Mass from a Plane. Let DE be the line in which any plane is cut by a perpendicular plane through P_1P_2 of the last section. Draw P_1L_1, P_2L_2, C_1M_1 perpendicular to DE and denote their lengths by d_1, d_2 and δ_1, respectively. Draw C_1A perpendicular to P_1L_1 and P_2B perpendicular to C_1M_1. Then $P_1A = (d_1-\delta_1)$ and $C_1B = (\delta_1 - d_2)$. From the similar triangles C_1AP_1 and P_2BC_1 we get

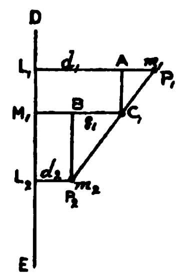


FIG. 37.

$$\frac{P_1A}{C_1B} = \frac{C_1P_1}{C_1P_2}$$

Hence

$$\frac{d_1-\delta_1}{\delta_1-d_2} = \frac{m_2}{m_1}$$

and $(m_1+m_2)\delta_1 = m_1d_1 + m_2d_2$.

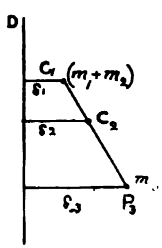


FIG. 38.

Hence *the product of the total mass by the distance of the center of mass from any plane equals the sum of the products of each mass by its distance from the plane.*

If we proceed to apply the same method to (m_1+m_2) at C_1 and m_3 at P_3 (Fig. 38) we must get a similar result expressed by the same rule. Hence

$$(\overline{m_1+m_2}+m_3)\delta_2 = (m_1+m_2)\delta_1 + m_3d_3$$

Hence substituting from the above

$$(m_1 + m_2 + m_3)\delta_3 = m_1d_1 + m_2d_2 + m_3d_3$$

By extending the same method to any number of particles we see that the law stated above (in italics) applies to the center of mass in all cases. This result might, in fact, have been taken as the definition of the center of mass.

It is evident that this result will not be altered if the order in which the various particles are taken be altered in any way. Hence the definition of the center of mass is not ambiguous.

For simplicity we have supposed that all the particles lie on one side of the plane considered, so that all the d's and δ's are positive. If in any case one or more of the distances are measured on the opposite side of the plane from the others, when we substitute numbers for the various distances those corresponding to one side of the plane must be given positive signs and the others negative. This can be shown by drawing the plane DE in Fig. 37 so that it cuts P_1P_r.

If the plane from which $d_1, d_2 \cdots$ are measured passes through the center of mass, δ is zero and in this case

$$m_1d_1 + m_2d_2 + \cdots = 0$$

76. Coördinates of the Center of Mass. When in any case it is desired to find the position of the center of mass of a body by applying the rule stated (in italics) in § 75, it is only necessary to apply the rule to distances from three planes at right angles. Denoting the distances from one of them by x's, from a second by y's, from the third by z's, we get

$$\bar{x} = \frac{m_1x_1 + m_2x_2 + \cdots}{M}, \quad \bar{y} = \frac{m_1y_1 + m_2y_2 + \cdots}{M},$$

$$\bar{z} = \frac{m_1z_1 + m_2z_2 + \cdots}{M}$$

where $\bar{x}$ denotes the distance of the center of mass from the plane from which the x's are measured, and similarly for $\bar{y}$ and $\bar{z}$.

77. Center of Mass of a Regular Body. The center of mass of two equal particles is at the middle of the line joining them. A

uniform rod may be divided into pairs of equal particles, the two in each pair being equidistant from the center of the rod. Hence the center of mass of the whole rod is at its middle point. Similar reasoning may be applied to any homogeneous body which has a geometrical center such as a circle, ellipse, sphere, spheroid, parallelogram, cube, parallelopiped, etc. The center of mass of each of these is at its geometrical center.

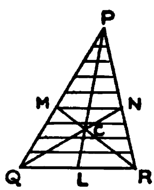


FIG. 39.

When a body can be divided into parts such that the center of mass of each is known, the center of mass of the whole can usually be found. A triangle may be divided into narrow strips parallel to one side; the center of mass of each strip lies on the line joining the middle of that side to the opposite vertex. Hence the center of mass of a triangle is at the intersection of the three lines which join the vertices to the middle of the opposite sides. Similar reasoning shows that the center of mass of a triangular pyramid is at the intersection of the four lines that join the vertices to the respective centers of mass of the opposite faces.

78. Velocity and Acceleration of the Center of Mass. Let us suppose that the velocity of each particle in a group of particles is known. How can the velocity of the center of mass be found? To answer this it is sufficient to show how the velocity of the center of mass in a direction perpendicular to each of three planes at right angles can be found.

To find the velocity of the center of mass in a direction perpendicular to any plane, consider the distances of the particles and of the center of mass from that plane. These are connected by the equation (§ 75)

$$(m_1 + m_2 + \cdots)\delta = m_1 d_1 + m_2 d_2 + \cdots \qquad (1)$$

At a time t later these distances will have all changed. Let the new values of the distances be d_1', $d_2' \cdots \delta'$. Then

$$(m_1 + m_2 + \cdots)\delta' = m_1 d_1' + m_2 d_2' + \cdots \qquad (2)$$

Subtract each side of (1) from the corresponding side of (2);

divide through by t and suppose t decreased without limit. Then $(\delta' - \delta)/t$ will become the velocity, say $\bar{v}$, of the center of mass; $(d_1' - d_1)/t$ will become the velocity, say v_1, of m_1 and so on. Hence

$$(m_1 + m_2 + \cdots)\bar{v} = m_1v_1 + m_2v_2 + \cdots \tag{3}$$

Thus the velocity of the center of mass is related to the velocities of the separate particles as the distance of the center of mass from any plane is related to the distances of the particles from that plane.

We may now proceed to apply the same reasoning to find the acceleration of the center of mass. Starting with (3) above let us consider what (3) becomes at a short time t later. We shall thus get two equations. Subtracting one from the other as before, dividing by t, and then supposing t indefinitely short, we get

$$(m_1 + m_2 + \cdots)\bar{a} = m_1a_1 + m_2a_2 + \cdots \tag{4}$$

Equation (3) is readily obtained by differentiating (1) with reference to the time (see § 19) and (4) is obtained by differentiating (3) (see § 31).

79. Acceleration of Center of Mass due to External Forces. Equation (4) of the last section has a very important interpretation. The term m_1a_1 is, by the Second Law of Motion, equal to the force that acts on m_1 in the direction in which a is measured, which of course may be any direction, and similarly for the other particles. Now the forces may be divided into two groups, (1) forces applied from the outside or *external* forces such as gravity acting on the body, pressures and pulls applied to the surface of the body and so on; (2) forces that the particles exert on one another, that is *internal forces*, actions and reactions between the particles. By the Third Law of Motion these internal forces occur in pairs of equal and opposite forces, and the sum of the components of all of them in any direction is zero.

Hence the right hand side of (4) stands for the sum of the components, in the direction considered, of all the external forces. Thus if M be the whole mass of the body or group of particles

$$a = \frac{\textit{sum of components of external forces}}{M}$$

Now by the Second Law of Motion this is the expression we would

arrive at if we asked, "what acceleration would the center of mass of the body receive if the whole mass were concentrated there and all the external forces were transferred parallel to themselves so as to act at that point?"

Hence *the center of mass of a body moves as if the whole mass were concentrated at the center of mass and the forces acting on the body were transferred, with their directions unchanged, to the center of mass.*

We now see the explanation of the facts stated in § 73. In the case of a body tossed into the air gravity is the only external force, and the center of mass moves as if all the mass and weight were concentrated there, that is, it moves as a particle would. Even when a body has its form changed very abruptly by the action of internal forces, as in the case of the explosion of a rocket, the internal forces do not affect the motion of the center of mass of all the particles. When two bodies approach and impinge, the motion of their center of mass is not affected by the forces between the bodies during impact, and hence continues unchanged after the impact. There are powerful forces of attraction between the sun and the planets that make up our solar system, but the center of mass of the whole moves with a uniform velocity through space.

80. Translation and Rotation. The preceding principles enable us to calculate the motion of the center of mass of a body, given the mass, the center of mass, and the external forces acting on the body. Now if we know the motion of the center of mass and the angular velocity of the body about an axis through the center of mass, we know the whole motion of the body. In the next chapter we shall see how to find the effect of a force that acts on a body so as to produce rotation.

MOMENTS OF FORCE AND MOMENTS OF INERTIA.

81. Thus far we have considered the effect of a force in producing motion of translation of a particle. When a force acts on an extended body (as distinguished from a body so small that it may be considered as a particle), the force will in general produce rotation as well as translation.

We shall first suppose that the effect produced by the force is one of rotation only, that is, we shall suppose that an axis in the

body is fixed so that translation is impossible. To find to what extent a force will produce rotation we must consider, as regards the force, something in addition to its magnitude and direction. For it is a matter of common experience that a force can be most effectively employed to set a heavy body, such as a fly wheel, in rotation when the force is applied as far as possible from the axis.

On the other hand when a force is applied to set such a body in rotation the inertia resistance it meets depends on something more than the mass of the wheel. For it is a matter of common experience that the farther on the whole the mass of the wheel is from the axis, *e. g.*, when the wheel has a heavy rim with a light hub and light spokes, the harder it is to set in motion or to stop. We are thus led to consider moments of forces and moments of inertia.

82. Moment of Force and Moment of Inertia. As the simplest case consider a particle P of mass m attached to an axis so that it is only free to move in a circle about this axis. Let a force F act on the particle. If the line in which the force acts is not in the plane of the circle, that is, if it is not perpendicular to the axis, we may resolve the force into two components, one parallel to the axis and one, f, perpendicular to the axis. The first component cannot affect the motion of the particle, since it acts in a direction in which the particle is not free to move. The second component, f, may again be divided into two components, one in the direction of the radius and the other along the tangent. The former of these cannot affect the motion of the particle but the latter, which is equal to $f \cos \theta$, where θ is the angle that f makes with the tangent, is in the direction in which the particle is free to move. Hence it will produce an acceleration a along the tangent and

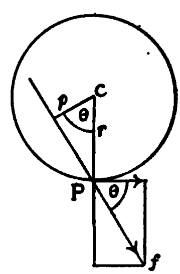


FIG. 40.

$$f \cos \theta = ma$$

Since this is a case of rotation only, the motion is more appropriately described by means of the angular motion about the center. Now the angular acceleration, α, is connected with the linear ac-

celeration by the relation (§ 69)

$$\alpha = ar$$

r being the radius of the circle. From the center of the circle drop a perpendicular p on the line of action of f. Then, since the angle which p makes with the radius through the particle is also θ, $\cos\theta = p/r$. Substituting these values we get

$$fp = mr^2 a$$

The product fp, that is the product of p by the component of F perpendicular to the axis, is called the ***moment of F*** about the axis, and mr^2 is called the ***moment of inertia*** of m about the axis. It should be carefully noted that in finding the moment of a force about an axis we resolve the force into two components, one of which is in the direction of the axis while the other is in a plane perpendicular to this axis, and it is on this latter component only that the moment of the whole force depends.

If two or more forces act on the particle, we would in the above have to take the component of each in a plane perpendicular to the axis and then again resolve these in the direction of the tangent. Thus we would get two or more terms, such as fp and $f'p'$, and the total moment of force acting on the particle would be the sum of these.

Since one direction of rotation about an axis is taken as positive while the opposite is taken as negative, moments of forces in the former direction must be considered as positive and in the latter as negative.

Let us now consider an extended body mounted on an axis and acted on by forces applied at various points. All particles of the body rotate with the same angular acceleration. On each particle, therefore, one or more forces act. We may, then, for each particle, write down an equation like the above. Let us now suppose the corresponding sides of all these equations to be added up. Then

$$f_1p_1 + f_2p_2 + \cdots = (m_1r_1^2 + m_2r_2^2 + \cdots)a$$

On the left-hand side we have the sum of the moments of all the forces acting on all the particles. Some of these are the external forces applied to the body, others are the internal forces, actions

and reactions, between the particles. Now by the Third Law of Motion these internal forces occur in pairs of equal and opposite forces, and the sum of the moments of such a pair is zero. Hence in finding the sum on the left side of the above equation we may omit all internal forces. Hence the left-hand side is the sum of the moments of the *external* forces applied to the body. This we call ***the total moment of force acting on the body*** and denote it by L. The multiplier of a on the right-hand side is the sum of the moments of inertia of all the particles of the body about the axis. This we call the ***moment of inertia of the body*** about that axis and denote by I. Hence

$$L = Ia$$

83. Moments of Inertia About Parallel Axes. The following general proposition is often of great assistance in finding the moment of inertia of a body.

If I_0 is the moment of inertia of a body of mass M about an axis through the center of mass, the moment of inertia, I, about a parallel axis at a distance h from the first axis is

$$I = I_0 + Mh^2$$

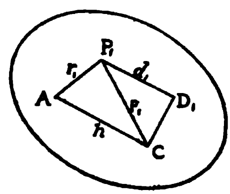


FIG. 41.

Consider a particle m_1 at P_1. From P_1 draw a perpendicular P_1A to the axis A and denote P_1A by r_1. Also draw a perpendicular P_1C to the parallel axis through the center of mass C and denote it by ρ_1. Suppose the same done for the particles m_2, m_3, $\cdots$. Denote the angle P_1CA by θ_1, P_2CA by θ_2 and so on. Then

$$\begin{aligned} I &= m_1r_1^2 + m_2r_2^2 + \cdots \\ &= m_1(\rho_1^2 + h^2 - 2\rho_1 h \cos\theta_1) + m_2(\rho_2^2 + h^2 - 2\rho_2 h \cos\theta_2) + \cdots \\ &= (m_1\rho_1^2 + m_2\rho_2^2 + \cdots) + (m_1 + m_2 + \cdots)h^2 \\ &\qquad\qquad - 2h(m_1\rho_1\cos\theta_1 + m_2\rho_2\cos\theta_2 + \cdots). \end{aligned}$$

Of these three terms which make up I the first two are I_0 and Mh^2 respectively. We shall show that the third term is zero. Draw through C a plane perpendicular to AC and denote the perpendicular P_1D_1 on the plane by d_1 and so on for the other particles. Then $\rho_1 \cos\theta_1 = d_1 \cdots$. Hence the third term above equals

$$2h(m_1d_1 + m_2d_2 + \cdots).$$

The expression in brackets is zero since d_1, d_2, $\cdots$ are the distances of the particles of the body from a plane through the center of mass (§ 75).

84. Two Fundamental Moments of Inertia. The finding of formulas for the calculation of moments of inertia usually requires the use of the integral calculus; but from a few fundamental cases many others can be deduced.

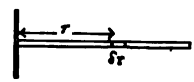


FIG. 42.

1. *The Moment of Inertia of a Thin Uniform Rod.* If the mass of the rod is M and its length L its moment of inertia about a transverse axis through one end is $I = \frac{1}{3}ML^2$. To find its moment of inertia, I_0, about a parallel axis through the center we must (by § 83) subtract $M(L/2)^2$ from I. Hence $I_0 = \frac{1}{12}ML^2$.

Let ρ be the mass of unit length of the rod. Consider an infinitesimal length dr at a distance r from one end; its moment of inertia about the transverse axis through that end is $\rho dr \cdot r^2$. Summing this up for all particles of the rod by the method of the Integral Calculus gives $\frac{1}{3}\rho L^3$ and substituting M/L for ρ we get $\frac{1}{3}ML^2$.

2. *Moment of Inertia of a Uniform Circular Disk.* Let the mass of the disk be M and its radius R. Its moment of inertia about a perpendicular axis through its center is $\frac{1}{2}MR^2$.

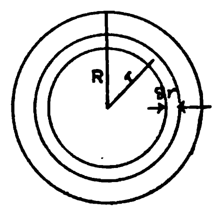


FIG. 43.

Let ρ be the mass per unit area of the disk. Consider a ring of mean radius r and width dr. Its area is $2\pi r dr$, its mass is $2\pi r dr \rho$ and its moment of inertia about the axis is $2\pi r^3 dr \rho$. Summing this up for all values of r from o to R gives $\frac{1}{2}\pi R^4\rho$, and, since the mass of the disk is $\pi R^2\rho$, the moment of inertia is $\frac{1}{2}MR^2$.

85. Moments of Inertia of a Disk. If the moments of inertia of a uniform disk about two perpendicular axes in the plane of the disk are I_1 and I_2 its moment of inertia I about a third axis intersecting the two and perpendicular to the disk equals $(I_1 + I_2)$.

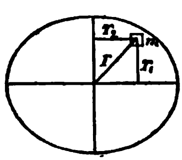


FIG. 44.

For suppose the disk divided into a very large number of very small parts and let m be one and suppose its distances from the two axes in the plane of the disk to be r_1 and r_2 and its distance from the third axis to be r. Then

$$r^2 = r_1^2 + r_2^2.$$

$$mr^2 = mr_1^2 + mr_2^2.$$

Hence

Summing up for all the particles,

$$I = I_1 + I_2.$$

From this and § 84 it follows that the moment of inertia of a circular disk about a diameter equals $\frac{1}{4}MR^2$.

86. Moments of Inertia of a Rectangular Disk. 1. *About an Axis in the Plane of the Disk and bisecting the pair of sides of Length a.* Suppose the disk divided into strips parallel to the a-sides. Applying to each the formula for a rod (§ 84) and adding for all the strips we get

$$I = M\frac{a^2}{12}$$

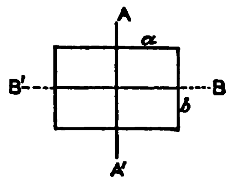


FIG. 45.

2. *About an Axis in the Plane of the Disk and Bisecting the Sides of Length b.*

$$I = M\frac{b^2}{12}$$

3. *About an Axis through the Center and Perpendicular to the Disk.* By § 85

$$I = I_1 + I_2$$
$$= M\frac{a^2 + b^2}{12}$$

87. Moment of Inertia of a Rectangular Block. A rectangular block may be regarded as made up of disks parallel to one face. If the sides of that face are of lengths a and b, the formula just stated will give the moment of inertia of each disk about an axis through the center and perpendicular to the face. Summing up both sides for all the disks and taking M for the whole mass and I for the whole moment of inertia we get

$$I = M\frac{a^2 + b^2}{12}$$

88. Moment of Inertia of a Circular Cylinder about its Geometrical Axis. Let the radius of the cylinder be R and its mass M. The cylinder may be divided into circular disks each of mass m by transverse sections. The moment of inertia of each is $\frac{1}{2}mR^2$. Summing the expression for all the disks we get.

$$I = \tfrac{1}{2}MR^2$$

About a Transverse Axis through the Center. Let the cylinder be supposed divided into disks as before. The moment of inertia of one of these about a diameter is $\frac{1}{4}mR^2$ (§ 85), and, if its distance from the center of the cylinder is x, its moment of inertia about the transverse axis through the center of the cylinder is $\frac{1}{4}mR^2 + mx^2$ (by § 83). This expression is to be summed for all the disks. The sum of the first term is $\frac{1}{4}MR^2$. The summation of the second term is the same as finding the moment of inertia of a thin rod of mass M and length L by dividing it into parts each of

mass m. Hence (§ 84) the sum of the second term is $\frac{1}{12}ML^2$ where L is the length of the cylinder. Therefore

$$I = M(\tfrac{1}{4}R^2 + \tfrac{1}{12}L^2).$$

89. Radius of Gyration. If the moment of inertia of a body about a certain axis is I and the mass of the body is M, and if we take a length k such that $I = Mk^2$, k is called the radius of gyration of the body about that axis. From the results of preceding sections we have: (a) for a uniform rod about its center $k^2 = \frac{1}{12}L^2$ and about one end $k^2 = \frac{1}{3}L^2$; (b) for a circular disk or cylinder about its axis $k^2 = \frac{1}{2}R^2$, and so on.

From the definition of the radius of gyration it follows that k is the distance from the axis at which we might suppose the whole mass of the body concentrated without changing the moment of inertia about that axis.

TABLE OF MOMENTS OF INERTIA.

Body.	Axis.	Moment of Inertia.
Rod	transverse through end	$\frac{1}{3}ML^2$
Rod	transverse through middle	$\frac{1}{12}ML^2$
Circular disk	perpendicular through center	$\frac{1}{2}MR^2$
Circular cylinder	longitudinal through center	$\frac{1}{2}MR^2$
Circular cylinder	transverse through center	$M(\frac{1}{4}R^2 + \frac{1}{12}L^2)$
Rectangular block	through center perpendicular to face with sides a and b in length	$\frac{1}{12}M(a^2 + b^2)$
Sphere	through center	$\frac{2}{5}MR^2$

90. Angular Momentum. Having considered the methods of calculating the moment of inertia of a body, we shall now consider further the equation between I, L and a for a body mounted on an axis (§ 82). Multiplying both sides of that equation by the time, t, during which the moment of force acts on the body, we get

$$Lt = Iat$$

Now at equals the increase of angular velocity. Hence if the angular velocity at the beginning of t is ω and at the end ω'

$$Lt = I(\omega' - \omega)$$

The product, $I\omega$. of the moment of inertia of a body about an axis and its angular velocity about the axis is called the *angular*

momentum (or moment of momentum) of the body about that axis. It corresponds to linear momentum, mv, in the case of translation, moment of inertia taking the place of mass and angular velocity the place of linear velocity. Lt, the product of a moment of force by its time of action, corresponds to the impulse of a force (§ 45); it is equal to the angular momentum produced in time t.

Since I is a constant for a given body and a given axis, any change in $I\omega$ is due to a change in ω, and the rate of change of $I\omega$ is I multiplied by the rate of change of ω, which is α. Hence $I\alpha$ is the rate of change of angular momentum, and, since it is equal to the moment of force, L, which produces it, *a moment of force about an axis equals the rate at which it produces angular momentum about that axis.* If L is zero, that is, if the total moment of force about an axis is zero, the angular momentum about the axis is a constant.

91. Conservation of Angular Momentum. The last statement is a particular case of the principle called the *Conservation of Angular Momentum,* namely, the case in which the body is a single rigid body attached to a fixed axis. The general principle is that *the total angular momentum of any system of bodies which is free from external influences remains constant.*

92. Kinetic Energy of Rotation. Each particle of a rotating body has a certain linear velocity and a certain amount of kinetic energy, and the total kinetic energy is the sum of the kinetic energies of all the particles. A particle of mass m at a distance r from the axis of rotation has a linear velocity ωr, and its kinetic energy is therefore $\frac{1}{2}m(\omega r)^2$. To find the total kinetic energy E we must sum up for all the particles. Now r is different for different particles but ω is the same for all. Hence

$$\begin{aligned} E &= \tfrac{1}{2}\omega^2(m_1r_1^2 + m_2r_2^2 + \cdots) \\ &= \tfrac{1}{2}I\omega^2 \end{aligned}$$

I being the moment of inertia of the body about the axis of rotation. This formula for kinetic energy of rotation is similar to the formula for kinetic energy of translation $\frac{1}{2}mv^2$, I taking the place of m and ω that of v.

93. Work Performed by the Moment of a Force. When a force

applied to a body rotates the body through an angle, the moment of the force about the axis of rotation does a certain amount of work. To find the amount let A be the axis of rotation (perpendicular to the paper) and f the component of the force in a plane perpendicular to A, and let the perpendicular AP be denoted by p. Then the moment, L, of the force about A is equal to fp. Suppose now that the body turn through an angle θ so that P comes to P', the force f remaining perpendicular to AP'. In this motion the force f has acted through the distance PP' along the arc of a circle, and, since PP' is equal to $p\theta$, the work done by the force is $fp\theta$ or $L\theta$.

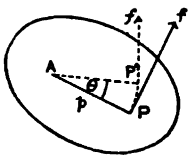


FIG. 46.

Hence *the work done by the moment of a force is the product of the moment by the angular displacement,* just as the work done by a force is the product of the force and the linear displacement.

If the motion of the body about A be not resisted by any other force the moment of force L will produce an angular acceleration a such that $a = L/I$, I being the moment of inertia about A. If, at the beginning of the angular displacement θ, the angular velocity is ω_0, and at the end of the displacement it is ω, by § 67

$$\omega^2 = \omega_0^2 + 2a\theta$$

Substituting the value of a we get

$$L\theta = \tfrac{1}{2}I\omega^2 - \tfrac{1}{2}I\omega_0^2$$

Hence in this case the work done by the moment of the force is equal to the increase of kinetic energy of rotation which it produces. This is only a particular case of the conservation of energy and might have been deduced at once from that principle.

If the work is done in twisting a spring, as in winding up a spring clock, the potential energy of the spring will be increased by the work so done.

94. Kinetic Energy of a Body which has both Translation and Rotation. For simplicity, in treating of the kinetic energy of a body we have considered the case of translation and that of rotation separately. But it frequently happens that a body has both motions simultaneously, *e. g.*, a body thrown at random into

the air or a body rolling down an incline. We have already seen (§ 80) that the whole motion of such a body may be regarded as consisting of two separate motions, a motion of translation of the center of mass and a motion of rotation about the center of mass, and these may be calculated separately. It can be shown mathematically (Duff's *Mechanics*, § 107) that we may proceed in the same way in finding the total kinetic energy of such a body. *The total kinetic energy of a body is the sum of the kinetic energy due to motion of translation of the center of mass and the kinetic energy due to motion of rotation about the center of mass.* From this we can readily calculate the total kinetic energy of a locomotive wheel, of a body rolling down an incline, and so on.

Because motion of rotation about the center of mass and motion of translation of the center of mass are independent as regards both acceleration and kinetic energy, motion of rotation about the *center of mass* is often called *pure rotation.*

RESULTANT OF FORCES ACTING ON A BODY.

95. Resultant. When treating of the forces acting on a *particle* we found that they could always be replaced by a single equivalent force called their resultant. When a number of forces act on a *body,* they are in certain cases equivalent in their effects to a single force, which is called their resultant. As we shall see later, there are other cases in which this is not so.

96. Conditions to be Satisfied by Resultant. 1. The resultant must be competent to produce the actual linear acceleration of the center of mass C, and, therefore, its component in any direction must equal the sum of the components of the acting forces in that direction. This condition is simplified by considering that any actual acceleration of C is made up of three independent components along axes at right angles. Hence *the resultant must have a component in each of three rectangular directions equal to the sum of the components of the forces in these directions.*

2. The resultant must be competent to produce the actual angular acceleration about any axis, and, therefore, its moment about any axis must equal the sum of the moments of the acting forces about that axis. It is, however, not necessary to consider all axes; for the whole motion of a body may be considered as made up of trans-

lation of the center of mass and rotation about the center of mass. Hence it is sufficient to consider axes through the center of mass. Whatever the angular acceleration about the center of mass it is equivalent to three component angular accelerations about rectangular axes through the center of mass (§ 72). Hence the second condition reduces to this: ***the moment of the resultant about each of three rectangular axes through the center of mass must equal the sum of the moments of the forces about that axis.***

If a force satisfies the above conditions it is the resultant. We shall now apply these tests to find the resultant of the forces acting on a body in some cases of importance.

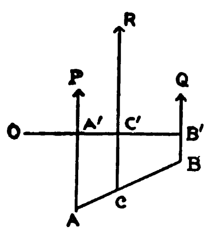


FIG. 47.

97. Resultant of Two Parallel Forces. 1. *Let P and Q be two forces in the same direction* acting at points A and B respectively. A single force R in the direction of P and Q and equal to $(P+Q)$ will satisfy the first condition of § 96, since its component in any direction equals the sum of the components of P and Q in that direction.

This force will also satisfy the second condition provided it act at a point C in AB such that

$$\frac{P}{Q}=\frac{CB}{CA}$$

For first consider an axis through the center of mass and perpendicular to the plane of P and Q. Suppose it to cut that plane in O. Draw $OA'C'B'$ to cut the lines of the forces at right angles. Then

$$\frac{P}{Q}=\frac{CB}{CA}=\frac{C'B'}{C'A'}=\frac{OB'-OC'}{OC'-OA'}$$

$$\therefore\ (P+Q)OC'=P\cdot OA'+Q\cdot OB'$$

Thus the moment of R about the axis equals the sum of the moments of P and Q. Next take an axis through the center of mass perpendicular to the above axis and to the lines of the forces. All the forces are at the same distance from this axis, and, since R equals $(P+Q)$, the moment of R about it equals the sum of the moments of P and Q about it. Finally an axis perpendicular

to the other two will be parallel to P, Q, and R and each will have zero moment about it.

Hence R is the resultant of P and Q.

It is important to notice that C is the point we would have found if we had been seeking the center of mass of particles at A and B proportional to P and Q (§ 74).

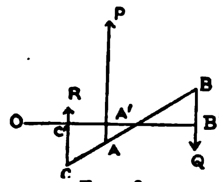


FIG. 48.

2. *Let P and Q be in opposite directions* and suppose $P>Q$. A single force R in the direction of the greater, P, and equal to $(P-Q)$ will satisfy the first condition, since its component in any direction equals the sum of the components of P and Q in that direction. It will also satisfy the second condition if it act at a point C in BA produced such that

$$\frac{P}{Q}=\frac{CB}{CA}$$

Consider first an axis through the center of mass perpendicular to the plane of the forces and cutting that plane in O and draw $OC'A'B'$ to cut the lines of the forces at right angles. Then

$$\frac{P}{Q}=\frac{CB}{CA}=\frac{C'B'}{C'A'}=\frac{OB'-OC'}{OA'-OC'}$$

$$\therefore\quad (P-Q)OC'=P\cdot OA'-Q\cdot OB'$$

Hence the moment of R about the axis equals the (algebraic) sum of the moments of P and Q. The same is true of two other axes taken as in case (1); the proof need not be repeated as it is identical with that there given.

Hence R is the resultant of P and Q.

To find the distance of C from A replace CB by $(CA+AB)$ in the above equation for the position of C. Then

$$CA=\frac{Q\cdot AB}{P-Q}$$

Hence CA is greater the less $(P-Q)$, that is, the more nearly the forces are equal.

98. Couples. Two equal and opposite forces, not in the same line, constitute a *couple*. If we attempted to find the resultant of two such forces by the method of the last section, it would give zero force at an infinite distance and such a force has no real existence. Hence a couple cannot be reduced to a single force.

The sum of the moments of two forces constituting a couple is the same about all axes perpendicular to the plane of the couple. For about an axis O between the forces the moments of the forces are in the same direction and their sum is $(P \cdot OA + P \cdot OB)$ or $P \cdot AB$; and about an axis O' not between the forces the moments are in opposite directions and the sum is $(P \cdot O'B' - P \cdot O'A')$ which again equals $P \cdot AB$.

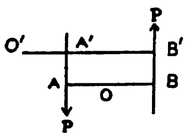


FIG. 49.

The distance AB between the forces of a couple is sometimes called the *arm* of a couple, and the moment of the couple about any axis perpendicular to its plane, that is $P \cdot AB$, is sometimes called the *strength* of the couple. Two couples in the same or parallel planes and of the same strength are equal in all respects and produce equal effects.

Since the sum of the forces of a couple equals zero, the couple produces no acceleration of the center of mass (§ 79); and if the center of mass be at rest it will remain at rest, or if it be moving in any way it will continue moving with constant velocity. The angular velocity produced by the couple must therefore be about some axis through the center of mass.

99. Resultant of any Number of Parallel Forces. To find the resultant of any number of parallel forces, whether in one plane or not, we may find the resultant of two, then combine this resultant with a third, and so on. The final resultant will be either a single force or a couple or zero. At each step the resultant equals the algebraic sum of the forces added. Hence the final resultant equals the algebraic sum of all the forces.

The line of action of the resultant may also be found by applying the principle that the moment of the resultant about any axis must equal the sum of the moments of the forces about that axis. When the forces are all in one plane, to find the line of action of the resultant we only need to take moments about any axis per-

pendicular to the plane. When the forces are not all in one plane it will be necessary to take moments about two rectangular axes perpendicular to the forces.

100. Center of Gravity. Attention has been called in (1) § 97 to the identity of the method of finding the resultant of parallel forces in the same direction and the method of finding the center of mass of a number of particles. If, for the particles in a certain group of particles or of a body, we substitute parallel forces all in one direction acting at the respective positions of the particles and proportional to the masses of the particles, the point of action of the resultant will coincide with the center of mass. This is sometimes taken as the definition of the center of mass. It should be noticed that nothing need be said as to the common direction of the parallel forces.

The forces of gravity on the particles of a body are (very nearly) parallel forces and they are proportional to the masses of the particles. Hence the *Center of Gravity* of a body, or the point of action of the resultant of the (very nearly) parallel forces of gravity, coincides with the center of mass of the body.

A very large body near the earth has a definite center of mass but not a definite center of gravity (except in some particular cases), for the forces are not quite parallel nor quite proportional to the masses. This is of no practical importance as regards bodies of the size found on the earth's surface; but it is of great importance in considering the effect of the attraction of the sun and moon on the motion of the earth.

101. Centrifugal Force. In § 47 we found an expression for the force required to keep a particle revolving in a circle. We may now extend this to *a body of any size or shape*. When a body of mass m rotates with constant angular velocity about any axis not through the center of mass, the latter moves uniformly in a circle and has therefore an acceleration v^2/r toward the center. Hence the force acting on the body (or the resultant of the forces if there are several) must, by the principle stated in § 79, equal mv^2/r and must act in the line joining the center of mass to the center of the circle and the body will react with an equal and opposite force. This reaction is the cause of the varying strain which an unbalanced fly-wheel exerts on the axis.

In many cases more than a single force (in addition to those required

to overcome friction and sustain the weight of the body) is required to keep a body rotating about an axis. As a simple case consider a pair of equal spheres joined by a light rod and rotating about a vertical axis through the center of the rod. Since the center of mass has no acceleration, the forces acting on the body if transferred to the center of mass would have a zero resultant. Hence the forces must form a couple and the reactions on the axis will form a couple, called a *centrifugal couple,* tending to bend the axis or make it rotate about an axis perpendicular to itself. For certain axes of rotation of a body the centrifugal couple is zero. In the above simple illustration this is true when the axis of rotation is in the line of the centers of the balls or at right angles thereto. These are also the positions of maximum and minimum moments of inertia of the body. A similar statement will evidently apply to a symmetrical body, such as a circular disk, which can be divided into pairs of particles like the above. Whatever the shape of a body there are three rectangular axes through any point of the body about which it can rotate without exerting any centrifugal couple.

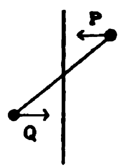


FIG. 50.

These are the axis of maximum moment of inertia through the point, that of minimum moment of inertia and a third perpendicular to both. These are called the *principal axes* through the point.

When a body is set spinning about a principal axis through its center of mass it continues to spin without any tendency to "wobble" or exert a centrifugal couple. This is illustrated by the motion of a well-thrown quoit or discus, by that of a bullet from a rifled gun and by the motion of the earth about its axis. But when the axis of initial spin is not a principal axis irregular motion ensues, as is illustrated by a badly thrown quoit.

FORCES IN EQUILIBRIUM.

102. Conditions of Equilibrium. The forces acting on a body are in equilibrium when they cause no acceleration either linear or angular, that is when their resultant is zero.

Given that a system of forces is in equilibrium we may conclude (from § 79) that *the sum of their components in any direction equals zero,* since there is no acceleration of the center of mass, and also that *the sum of their moments about any axis equals zero,* since there is no angular acceleration about any axis.

When we equate the sum of the components of the forces in any direction to zero we get a relation between the forces, and it might seem that we could get an unlimited number of such relations; but in reality there are only three of these independent, *e. g.,*

those got by taking the sum of the components in some three directions at right angles.

Similarly we get a relation between the forces by equating the sum of the moments about any axis to zero; but again there are only three of these relations independent, *e. g.*, those got by taking moments about some three rectangular axes.

Thus we can deduce at most six independent relations between forces in equilibrium and this might have been expected from the fact that a body has six degrees of freedom at most—three of translation and three of rotation.

We may reverse the point of view and ask what relations and how many must forces satisfy to make it certain that they shall be in equilibrium, that is, what are the conditions essential to equilibrium. The answer is again six relations, namely, the sum of the components in each of any three rectangular directions must equal zero and the sum of the moments about each of some three rectangular axes must equal zero.

103. Forces in a Plane. When the lines of action of forces that are in equilibrium lie in one plane, the sum of the components of the forces in each of any two directions at right angles in the plane equals zero. In this case the third rectangular axis is perpendicular to the plane and the component of each force in that direction is zero. Also the sum of the moments of the forces about any axis perpendicular to the plane is zero. The other two rectangular axes are in the plane and the moment of any one of the forces about such an axis is zero.

Hence when forces in a plane are in equilibrium three independent relations among the forces can be deduced.

104. Examples of Equilibrium of Forces in a Plane. To illustrate the above we shall consider two examples.

1. A uniform beam AB (length $= l$) rests without slipping on the ground and leans without friction against a smooth wall. What is the force (F_1) on the wall and on the ground (F_2) and what is the force of friction (F_3) between the beam and the ground?

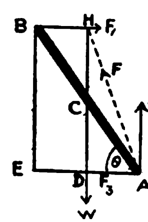


Fig. 51.

Since there is no friction at B, F_1 is horizontal. The force of friction at A, that is F_3, is horizontal and toward E. Equating the sum of the horizontal forces to zero we get

$$F_1 - F_3 = 0 \qquad (1)$$

and equating the vertical forces to zero we get

$$F_3 - W = 0. \tag{2}$$

A third relation may be obtained by taking moments about any axis perpendicular to the plane of the forces. If we choose for this purpose an axis through A, the relation will be as simple as possible, since F_2 and F_3 have zero moment about such an axis. The weight acts at the center C of the beam and the distance of its line of action from A is $(l/2)\cos\theta$. Also the distance BE of the line of action of F_1 from A equals $l\sin\theta$. Hence

$$W\frac{l}{2}\cos\theta - F_1 l\sin\theta = 0. \tag{3}$$

From these three equations we get

$$F_1 = F_2 = \tfrac{1}{2}W\cot\theta$$
$$F_3 = W.$$

2. A uniform rod hangs from a vertical wall by a hinge and rests on a smooth floor. In this case the force at A must be vertical, since there is no horizontal force of friction at A. Let the force on the beam at B consist of a horizontal part F_1 and a vertical part F_3. Equating to zero the sum of the vertical forces, the sum of the horizontal forces and the sum of the moments about B, we get

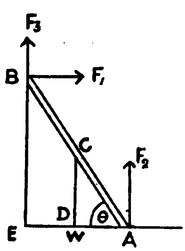


FIG. 52.

$$F_1 = 0;\ F_2 + F_3 - W = 0$$

$$W\frac{l}{2}\cos\theta - F_2 l\cos\theta = 0.$$

Hence $\quad F_1 = 0;\ F_2 = \tfrac{1}{2}W;\ F_3 = \tfrac{1}{2}W.$

Since F_1 is zero the rod does not press against the wall. This result, which seems at first improbable, may be verified by allowing A to rest on a board on a tank of water and hanging B by a cord; the cord will be found to be vertical when tested by comparison with a plumb line.

105. Special Cases of Equilibrium. 1. *When two forces are in equilibrium they must be equal and opposite and in the same line.* If not equal and opposite they would produce translation and if not in the same line they would produce rotation.

For example a body suspended by a cord must rest so that its center of gravity is vertically below the point of support. This

supplies an experimental method of finding the center of gravity of a disk of any shape. It is only necessary to support it in succession at two points on its rim and find the intersection of the lines of support.

2. *When three forces are in equilibrium they must all lie in one plane.* For the sum of the moments of all three about any axis is zero. About any axis that intersects the lines of action of two of the forces the moments of these two forces are zero. Hence any such axis must also intersect the line of action of the third force (unless it be parallel to it). Thus an infinite number of straight lines can be drawn so as to intersect the lines of action of all the forces and this cannot be so unless all the forces be in one plane.

3. *Three forces in equilibrium must either be parallel or pass through a single point.* If they are parallel one is equal and opposite to the resultant of the other two. If they are not parallel two of them intersect and their moments about the point of intersection are zero. Hence the third must pass through the point of intersection of any two.

As an example of three parallel forces in equilibrium consider (2) of § 104. The resultant of F_2 and F_3 must be equal and opposite to and in same line as W which acts at the middle of AB. Hence F_2 and F_3 are equal.

As an example of three non-parallel forces in equilibrium consider (1) of § 104. Let the resultant of F_2 and F_3 be F. Then F, F_1 and W are three forces in equilibrium. Hence F must pass through the intersection of F_1 and W. Hence the direction of F is readily found graphically. We may also find graphically the magnitudes of F_1 and F. Since DA and BH are equal, $HBDA$ is a parallelogram. Hence F, F_1 and W are proportional to HA, HB and HD.

106. Stable, Unstable and Neutral Equilibrium. A body is in equilibrium when it is either at rest or moving uniformly, that is, without acceleration linear or angular. The resultant of the forces acting on such a body is zero.

When a body in equilibrium is at rest the equilibrium is described as *static.* Of this kind of equilibrium there are three forms, stable, unstable and neutral. A body at rest is in *stable equilibrium* when, on being slightly displaced, it tends to return to its equilibrium position. This is illustrated by a chemical balance,

a pendulum or picture hanging by a cord, a book on a table and in fact by most stationary objects. A body at rest is in *unstable equilibrium* when, on being slightly displaced, it tends to move further from its equilibrium position. An egg on end and a board balanced on one corner would be in unstable equilibrium. A body at rest is in *neutral equilibrium* when, on being slightly displaced, it has no tendency either to move further away or to return; for example, a sphere or cylinder on a horizontal table and any body suspended by an axis through its center of gravity.

A body in a position of stable equilibrium oscillates about that position when displaced and released, though the oscillation may

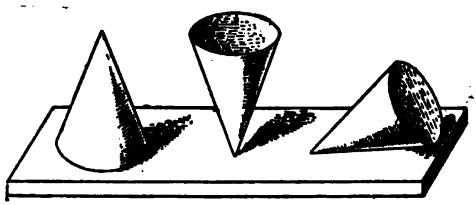


FIG. 53. Stable, unstable and neutral equilibrium.

be quickly destroyed by friction or other forces. When too far displaced such a body may come to a position of unstable equilibrium and not return; a table or chair tilted too far comes to a position of unstable equilibrium. The extent to which any such body may be displaced and yet return is a measure of the degree of stability of the equilibrium.

107. Energy Test of Static Equilibrium. When a body at rest is in stable equilibrium a disturbance will increase its potential energy. This is evident in the case of a pendulum at rest, for a disturbance raises its center of gravity; work is done against gravity when the body is displaced and this work produces potential energy. Thus *a position of stable equilibrium is a position in which the potential energy is a minimum.* This statement holds true whatever the force against which work is done; the fact that the disturbance produces an increase of potential energy shows that there are conservative forces opposing the motion and these forces will cause the body to return when it is displaced.

A position of unstable equilibrium is a position in which the potential energy is a maximum, as is illustrated by a spheroid on

end or a board balanced on a corner; a disturbance lowers the center of gravity. The statement is true whatever the forces in action; for the fact that the body when disturbed moves farther away from its position of equilibrium and thus gains kinetic energy shows that its potential energy diminishes.

When the equilibrium is neutral a displacement produces no change of potential energy; when a sphere rolls on a horizontal table its center neither rises nor falls. An interesting illustration is afforded by the apparatus sketched in the adjoining figure. It will remain at rest whatever the positions of the equal weights which are adjustable along the horizontal rods, for the total potential energy is the same in all.

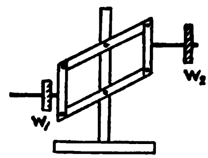


FIG. 54. Neutral equilibrium.

For *kinetic equilibrium* consult advanced works.

The principle that for stable equilibrium the potential energy is a minimum is extensively illustrated in nature; the potential energy may be partly or wholly other than mechanical energy, in forms dealt with in other parts of Physics. Changes are continually taking place in nature and bodies, when disturbed, settle into position of stable equilibrium, that is, of minimum potential energy.

KINEMATICS AND DYNAMICS

PERIODIC MOTIONS.

108. A **periodic motion** is one that is repeated in successive equal intervals of time. The time required for each such repetition is called the *period* of the motion. Thus the moon revolves around the earth with a periodic motion, the period of which is a lunar month and the earth revolves about the sun in a period of a year. The end of a hand of a clock has a periodic motion about the center of the face. A point on a vibrating violin string or piano wire has a periodic motion.

109. Uniform Circular Motion. When a point P revolves with constant speed in a circle of center C, the position of P at any moment may be assigned by giving the angle that CP makes with some fixed diameter such as CA. The angle is usually called the *phase* of *P's* motion.

If the period of the motion be T, the angle through which CP

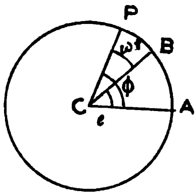


Fig. 55.

revolves in unit time is the angular velocity ω and equals $2\pi/T$. Let us suppose that at the moment from which we begin reckoning time P is at some position B and let its phase at that moment, that is, the angle BCA, be e. After time t, P will have revolved through an angle ωt or $(2\pi/T)t$ and the phase at time t will be $((2\pi/T)t + e)$. The angle e or the phase at zero time is called the *epoch* of the motion.

110. Simple Harmonic Motion. This is the most important form of periodic motion. *It is a vibration in a straight line, the motion being such that the vibrating point has an acceleration which is toward the center of its path and proportional to its distance from the center.*

Let $A'A$ be the path of vibration of a point M which has a simple harmonic motion, and let C be the center of $A'A$. Denote the distance of M from C at any time by x, and let values of x be considered as positive when M lies between C and A and negative when M lies between C and A'. When x is positive the acceleration, a, of M is toward C and is, therefore, in the negative direction, and when x is negative a, being still toward C, is positive. Hence, if we denote the constant of proportionality of the magnitude of a to x by c, by the above definition of simple harmonic motion

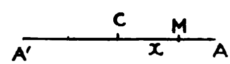


Fig. 56.

$$a = -cx$$

The distance x of the vibrating point from the center of motion is called the **displacement** *of the point.*

One-half of the length of the path of vibration is called the **amplitude** *of the simple harmonic motion.* We shall denote it by r. It is equal to the magnitude of the greatest displacement (CA or CA').

The time required for a complete vibration (that is from A to A' and back to A) is the **period** of the simple harmonic motion.

111. The Force Acting on a Body that has Simple Harmonic Motion. A body that has a simple harmonic motion has a varying

acceleration which is always directed toward the center. To produce this acceleration a varying force, also directed toward the center, must act on the body. Denote the force by F and let m be the mass of the body. From the Second Law of Motion and the definition of simple harmonic motion we get

$$\begin{aligned} F &= ma \\ &= -mcx \end{aligned}$$

Since m and c are constants for a given body and a given simple harmonic motion, *the force* required is always *opposite to and proportional to the displacement.*

The force required to stretch or compress a spiral spring one end of which is fixed is proportional to the displacement of the free end from its unstrained position, and the reaction exerted by the spring is opposite to and proportional to the displacement (§ 56). Hence a body attached to such a spring and allowed to vibrate under the action of the spring has simple harmonic motion. The same law of force holds for a flat spring when bent and, in fact, for any elastic body when distorted. Hence all elastic vibrations are simple harmonic motions or compounded of such motions and the same is true of the vibrations that constitute sound and light.

112. Energy of a Body that has Simple Harmonic Motion. The principle of the Conservation of Energy applies to a body that has simple harmonic motion, since the only force acting on the body is one that depends on the position of the body (§ 63). We have in fact already found in § 61 the proper expression for the total energy of such a body in any position; all we need to do is to substitute for k its value in the present case, namely mc. Hence the total energy is $(\frac{1}{2}mv^2 + \frac{1}{2}mcx^2)$, of which the first part is the kinetic energy at displacement x and the second is the potential energy. At one end of the path of vibration v is zero and $x = r$. Hence the total energy is potential and equal to $\frac{1}{2}mcr^2$. At the center x is zero and v has its largest value V; hence the energy is entirely kinetic and equal to $\frac{1}{2}mV^2$.

113. Velocity in Simple Harmonic Motion.—From the result just stated we can find a useful expression for the velocity at any

displacement. Equating the total energy at displacement x to that at maximum displacement we have

$$\tfrac{1}{2}mv^2 + \tfrac{1}{2}mcx^2 = \tfrac{1}{2}mcr^2$$
$$\therefore\ v = \pm\sqrt{c}\sqrt{r^2 - x^2}$$

and, referring to Fig. 56, it will be seen that the positive sign must be taken for motion from A' to A and the negative for motion from A to A'.

114. Simple Harmonic Motion may be Regarded as a Projection of Uniform Circular Motion. If the expressions for the acceleration and velocity of a point that has simple harmonic motion, namely,

$$a = -cx$$
$$v = \pm\sqrt{c}\sqrt{r^2 - x^2}$$

be compared with those for the projection of a uniform circular motion (§ 70), it will be seen that the two motions are of the same nature and *we may regard any simple harmonic motion as the projection of an imaginary uniform circular motion.* This, in fact, is sometimes taken as the definition of simple harmonic motion. The relation enables us to deduce in a simple manner some important properties of simple harmonic motion.

115. Period of a Simple Harmonic Motion. In the circular motion from which a given simple harmonic motion may be projected the radius of the circle must equal the amplitude of the simple harmonic motion. Moreover from the expression for a and v in § 113 and those in § 70 it is seen that the angular velocity ω in the circular motion must equal $\sqrt{c}$. Also the period of the circular motion and that of the simple harmonic motion must be equal.

The period of the circular motion equals $2\pi/\omega$ and this equals $2\pi/\sqrt{c}$ or $2\pi\sqrt{-(x/a)}$. Hence, if T is the period of the simple harmonic motion,

$$T = 2\pi\sqrt{-\frac{x}{a}}$$

It will be noted that x and a are always of opposite sign, hence the quantity under the radical is numerically positive.

116. Trigonometrical Expression for the Displacement. Let us consider more closely the circular motion of which a simple harmonic motion of amplitude r and period T may be imagined to be the projection. On the path $A'A$ of the simple harmonic motion as diameter describe the circle. Let P be the point whose motion projects into that of the vibrating point M. The angle PCA or the phase of P's motion equals $[(2\pi/T)t + e]$. Hence for CM or the displacement, x, we have

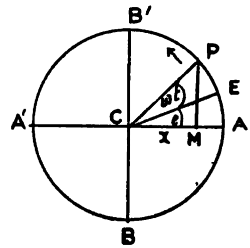


FIG. 57.

$$x = r \cos\left(\frac{2\pi}{T}t + e\right)$$

While we have deduced this expression from the related circular motion, it must now be regarded as simply an expression for the simple harmonic motion of amplitude r and period T. $[(2\pi/T)t + e]$ is called the *phase* of the simple harmonic motion at time t and e is called the *epoch* of the simple harmonic motion. Since the phase reduces to e when t is zero, the epoch is the phase at the moment from which time is reckoned.

For certain particular values of e the expression for x becomes simpler. If e is zero, which, as we see from the circular motion, means that at zero time M is at A, the expression for x is

$$x = r \cos\frac{2\pi}{T}t$$

If $e = -(\pi/2)$, at zero time P is at B and M is therefore at C and moving in the positive direction.

Now

$$\cos\left(\frac{2\pi}{T}t - \frac{\pi}{2}\right) = \sin\frac{2\pi}{T}t$$

Hence

$$x = r \sin\frac{2\pi}{T}t$$

117. Simple Pendulum. A simple pendulum consists of a small heavy body, called the bob (usually spherical), suspended by a

practically inextensible cord, the mass of which is so small as to be negligible compared with the bob. As the pendulum swings through a small angle the bob vibrates through a small arc of a circle which is very nearly a straight line.

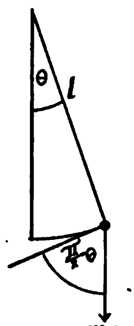


FIG. 58. Simple pendulum.

The force of gravity mg on the bob of the pendulum acts vertically and it may be resolved into a component along the tangent and a component along the radius. The latter component produces a tension on the cord which does not affect the motion, while the former component produces an acceleration along the tangent. When the cord is at an inclination θ to the vertical, the component along the tangent equals $mg \cos((\pi/2)-\theta)$ or $mg \sin \theta$. Since the pendulum is supposed to vibrate through a very small angle $\sin\theta$ may be replaced by θ; in fact, for values of θ less than $2°$, $\sin\theta$ and θ are equal within one part in 10,000. If the distance of the bob from its lowest point, measured along the tangent, be denoted by x and the length of the pendulum by l, $\theta = x/l$ radians. Hence the force along the tangent is $mg(x/l)$. This force is in the negative direction when x is positive. Hence, denoting the acceleration along the tangent by a, we have by the Second Law of Motion

$$-mg\frac{x}{l} = ma$$

Hence

$$a = -\frac{g}{l}x$$

Since the multiplier of x is a constant, the acceleration is opposite to and proportional to the displacement. Hence the motion is simple harmonic motion, and, if T be the period or time of vibration of the pendulum, by § 115

$$T = 2\pi\sqrt{-\frac{x}{a}}$$

$$= 2\pi\sqrt{\frac{l}{g}}$$

118. Angular Harmonic Motion. A body attached to an axis may vibrate backward and forward through an angle, as in the case of a balance wheel of a watch or of any heavy body hung on a peg. When the angular acceleration, α, is always opposite to and proportional to the angular displacement, θ, the motion is called ***angular harmonic motion.*** Hence the general formula for such motion is

$$\alpha = -C \cdot \theta$$

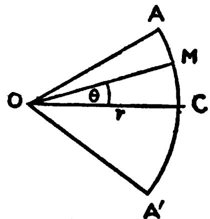


FIG. 59.

C being a constant.

Let Fig. 59 be a plane through the body perpendicular to the axis O. A line OM in the body will vibrate backward and forward through an angle. The point M will vibrate in an arc of a circle of radius OM or r. When the angular displacement of OM from its mean position OC is θ, the displacement, x, of M from C is $r\theta$ and the linear acceleration, a, of M is $r\alpha$ (§ 69). Substituting these values of θ and α in the above formula and concelling r we get

$$a = -C \cdot x$$

Thus the motion of M is simple harmonic motion in all respects except that it is along an arc (which may be long or short) instead of along a straight line. We might suppose the arc straightened out without any other change in the nature of the motion of M. Hence if T be the period of M's motion, which of course is the same as the period of the angular harmonic motion,

$$T = 2\pi\sqrt{-\frac{x}{a}}$$

$$= 2\pi\sqrt{-\frac{\theta}{\alpha}}$$

The expression for the calculation of the period of an angular harmonic motion is similar to that for the calculation of the period of a simple harmonic motion (§ 116).

As examples of angular harmonic motion we shall consider the torsion pendulum and the physical pendulum.

119. The Torsion Pendulum. A torsion pendulum consists of a vertical wire carrying a body at one end and clamped at the other end. When the body is turned around the wire as axis and released it performs angular vibrations; the twisted wire begins to untwist and thus starts the motion which persists after the wire has untwisted, owing to the kinetic energy acquired by the body.

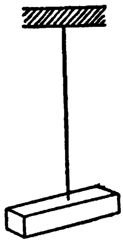


FIG. 60. Torsional pendulum.

To twist the wire requires the application of a couple. The twist θ produced by a certain couple of moment L is proportional to L and to the length l of the wire. Hence $Ll = \tau\theta$, where τ is a constant which is called the *constant of torsion* of the wire. The couple exerted by the twisted wire is equal and opposite to that required to produce the twist. Hence the couple exerted by the wire on the body is $-\tau(\theta/l)$ when the displacement is θ. This couple gives the body an angular acceleration, and, if we denote this by a and the moment of inertia of the body by I,

$$-\tau\frac{\theta}{l} = Ia$$

$$\therefore\ a = -\frac{\tau}{Il}\cdot\theta$$

In this the multiplier of θ is a constant which depends on the wire and the body and is independent of the motion. Hence the motion agrees with the definition of angular harmonic motion, and, if T is the period of vibration,

$$T = 2\pi\sqrt{-\frac{\theta}{a}}$$

$$= 2\pi\sqrt{\frac{Il}{\tau}}$$

It should be noticed that we have not assumed the angle of vibration to be small, as in the case of the ordinary pendulum; in the torsional pendulum the restoring couple is proportional to the angular displacement even when the latter is large (provided it is not so large as to permanently strain the wire).

By means of the torsional pendulum the moment of inertia of an irregular body can be compared with that of a body of known moment of

inertia. The two are, by the above formula for T, proportional to the squares of the corresponding times of vibration when the bodies are in turn attached to the same wire and set into angular vibration.

120. The Compound Pendulum. A body of any shape suspended by a horizontal axis and vibrating under gravity through a small angle constitutes a compound pendulum. Fig. 61 represents a vertical section through the center of gravity C and perpendicular to the axis of suspension S. Denote SC by h. When SC is inclined at an angle θ to the vertical, the force of gravity, mg, which acts at C, has a moment about S equal to $mgh \sin\theta$, which is negative when θ is positive. Hence if I is the moment of inertia of the body about S

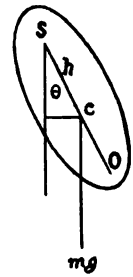


FIG. 61. Physical pendulum.

$$-mgh \sin\theta = I\alpha$$

If the angle θ is always small, we may, as in the case of the simple pendulum, replace $\sin\theta$ by θ and thus get

$$\alpha = -\frac{mgh}{I}\cdot\theta$$

This satisfies the condition for angular harmonic motion and the period of vibration is

$$T = 2\pi\sqrt{-\frac{\theta}{\alpha}}$$
$$= 2\pi\sqrt{\frac{I}{mgh}}$$

Let the radius of gyration about an axis through C parallel to the axis of suspension be k. Then the moment of inertia about the axis through C is mk^2 and about the parallel axis through S it is $mk^2 + mh^2$ (by § 83), which is, therefore, the value of I. Hence

$$T = 2\pi\sqrt{\frac{k^2 + h^2}{gh}}$$

If this be compared with the formula for a simple pendulum, it is seen that, if l be the length of a simple pendulum that vibrates in the same time as the compound pendulum,

$$l = \frac{k^2 + h^2}{h} = \frac{k^2}{h} + h.$$

Hence $$(l-h)h=k^2.$$

The length l is evidently greater than h. Hence if we measure along SC a length equal to l we shall arrive at a point O in SC extended. The point O, which is always on the opposite side of C from S, is the point at which the whole mass of the body might be supposed concentrated without any alteration of the period of vibration. O is called the **center of oscillation** corresponding to the axis of suspension S. Since $CO=(l-h)$ and $CS=h$, we have as the relation between any center of oscillation and the position of the corresponding axis of suspension

$$CS\cdot CO=k^2.$$

If the pendulum be now inverted and set to vibrate about an axis through O parallel to the former axis, the new center of oscillation O' will lie in OC produced and must satisfy the relation

$$CO\cdot CO'=k^2.$$

A comparison of these two equations show that O' must coincide with S. Hence the center of suspension and the center of oscillation are interchangeable and the distance between them is the length of the equivalent simple pendulum. This is the principle of Kater's pendulum.

121. Energy Changes. The resultant force of gravity acts at C (Fig. 61), hence the potential energy of the pendulum in any position is the same as if its mass were concentrated at C. But the pendulum does not swing as if it were concentrated at C, because its kinetic energy is that of its mass supposed concentrated at C plus its kinetic energy of rotation about C (§ 94). As the pendulum falls toward the vertical the lost potential energy goes partly into energy of rotation about C; hence it does not swing as rapidly as if it were concentrated at C, that is, as if it were a simple pendulum of length SC. A parallel case that brings out the distinction is illustrated by a block suspended by two cords as in Fig. 62. Swinging perpendicularly to the plane of the figure it is a physical pendulum of length SO, the block having energy of rotation. Swinging parallel to the plane of the figure it is a simple pendulum of length equal to the length of the cords; the block in this case has no rotation. A similar explanation applies to the motion of the pans of a balance. They do not rotate with the beam but move vertically; hence they affect the motion as if concentrated on the supporting knife-edges.

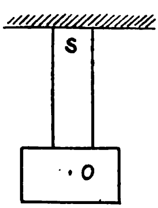


FIG. 62.

122. Center of Percussion. There is another important relation between an axis of suspension S and the corresponding center of oscillation O. A blow at O transverse to SO will start the body rotating about S without any jar on the support at S. Hence O is also called the center of percussion of the body when suspended at S. The center of percussion is readily found by suspending the body by a cord and striking horizontal blows at various points. Or it may be found by holding the body at S and striking across a table edge as a base-ball player strikes a ball with a bat; when the blow is through the center of percussion there is no jar on the hand.

FIG. 63. Center of percussion.

123. Gyroscopic Motion. A gyroscope is a wheel on a horizontal axle which is supported on a pivot (a bicycle wheel suspended by a vertical cord attached to a short extension of the axle will serve). When the wheel is set in rotation and the axle then released, the axle, instead of tilting in a vertical plane as it would if the wheel were at rest, revolves in a horizontal plane at a rate that depends on the velocity of rotation of the wheel about the axle. This motion is called **precession.** (Slight vertical oscillations or *nutations* of the free end of the axle may also accompany the precession.) The weight of the wheel acting at the center of the wheel has a moment about an axis through the pivot at right angles to the axis of the wheel. If this moment of force be increased by hanging a weight on the frame the rate of precession will be greater. If the wheel be supported at its center of gravity there will be no moment of force and no precession. (Thus mounted the instrument is sometimes called a gyrostat.)

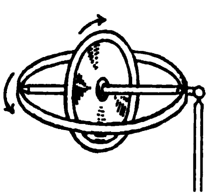


FIG. 64. A gyroscope.

If the motion be carefully considered it will be seen that it is very analogous to the revolution of a particle in a circle under the action of a force directed toward the center.

Circular Motion.	*Gyroscopic Motion.*
The force, F, is in a direction at right angles to the direction of motion.	The moment of force, L, is about an axis at right angles to the direction of the axis of rotation.

Circular Motion.	*Gyroscopic Motion.*
This force causes steady revolution of the particle with angular velocity ω.	This moment of force causes steady revolution of the axis of rotation with angular velocity ω'.
The momentum, mv, of the particle remains constant in magnitude but changes steadily in direction.	The angular momentum, $I\omega$, of the body remains constant in magnitude but the axis changes steadily in direction.

124. Moment of Force Required to Produce Precession. From the similarity of the two motions it might be expected that there would be a similarity in the formulas for the motions. The force in the circular motion is $m\ (v^2/r)$ or $mv \cdot \omega$, where ω is the angular velocity. Hence the force equals the product of the momentum mv and the rate of change of the direction of the momentum ω. Similarity the moment of force on the wheel equals the product of the angular momentum $(I\omega)$ of the wheel and the rate of change of the direction of the axis, *i. e.*, the rate of precession ω'. If L is the moment of the force $L = I\omega\omega'$

125. Other Examples of Precession. The curvature of the path of a coin rolled with a tilt along a table is due to the precession of its axis caused by the moment of its weight about the point of contact with the table. The motion of a top is a precession.

Any large body, such as a dynamo armature, in rotation aboard a vessel that is rolling, pitching, or turning, has a precessional motion and the bearings must supply the necessary moment of force and experience an equal and opposite reaction.

When a side-wheel steamer is turned in a sharp curve there is a precession of the axis of the paddle wheels. To produce this precession and at the same time keep the vessel level would require a moment of force about a longitudinal axis, and in the absence of such a moment the vessel lists to one side or the other.

The earth is not quite spherical but bulges at the equator. On one side the protuberance is closer to the moon than the center of the earth and on the other side it is farther away. The result of this (and of a similar but smaller moment exerted by the sun) is a moment of force that causes a precession of the earth's axis.

The gyroscope has been applied to steering torpedoes, to pre-

venting the rolling of ships, to balancing trains on a single rail, and it has been suggested as a means of balancing an aeroplane.

FRICTION.

126. Static Friction. When two solids are in contact there is a resistance, caused by the surfaces, to the sliding of one on the other. This resistance is called *Friction.* When a force parallel to the surfaces of contact is applied to one of the bodies and the force is less than a certain amount, which depends on the nature of the surfaces and the pressure between them, motion will not take place, the resistance being equal to the force. When the force is increased to a certain value the resistance will fail to increase and sliding will take place. This maximum resistance is called the ***maximum static friction.*** With a given pair of surfaces in contact and with a force tending to produce sliding motion in a certain direction (to avoid the influence of grain), ***the maximum static friction is found to be*** (within certain wide limits of pressure) ***proportional to the pressure.*** Denoting the coefficient of proportionality by μ, the maximum of static friction by F, and the pressure by P, we have

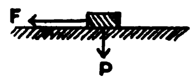


FIG. 65.

$$F = \mu P$$

The constant μ is called the ***coefficient of static friction,*** and may be defined as ***the ratio of the maximum static friction between two surfaces to the pressure between them.***

By the pressure here is meant the total perpendicular force between the surfaces (not the force per unit of area, as the word pressure is sometimes used). If one of the two bodies rests on the other and if the surfaces of contact are plane and horizontal, the pressure is the weight of the upper. The maximum static friction is the force applied horizontally to the upper that will just produce motion. If additional weights be placed on the upper body the pressure between the surfaces will be increased and the friction will be increased in the same proportion. If the upper

mass be redistributed in any way, for instance if it be cut in two and one part placed on the other, the total force of friction will not change; for, while the area of contact will be diminished, the pressure on each unit of area will be increased in the same proportion. Thus *the total frictional resistance is independent of the area of contact* and for two given surfaces depends only on the pressure, as is implied in the equation $F = \mu P$.

The coefficient of static friction between two surfaces depends on the materials and a variety of circumstances. The rougher the surfaces, that is the greater the inequalities in each, the larger is μ. If the surfaces are not clean parts of the surfaces are replaced by surfaces of the foreign substance and μ is necessarily different. The longer two surfaces are in contact the greater the maximum static friction; this is especially true of soft or fibrous surfaces. When the materials are of grained structure the friction is greater across the grain than along it. Friction is no doubt due to interlocking of the projections on one surface with those on the other surface. When slipping takes place some projecting pieces are broken off or abraded as it is called. With prolonged contact between two surfaces small readjustments of the surface particles take place, so that the fit becomes closer and the resistance to motion greater. It has even been found that when one surface is pushed a very small distance (0.01 of a cm.) it will when released spring back, thus showing that there is some elastic bending of surface projections. In general, friction between two surfaces of the same material is greater than between surfaces of different material. Thus there is an advantage in using brass bearings for steel shafts to diminish friction and covering with leather the face of a pulley used with leather belting increases friction and helps to prevent slip.

Friction is utilised in the transmission of energy by machine belting. Usually some slipping takes place for the belt stretches somewhat while in contact with the pulley. Friction between the driving wheels of a locomotive and the rails prevents slipping; without it the locomotive would be helpless, and where it is not sufficient the track is sanded. To hold a rope fast it is sometimes wrapped around a post. The friction on each part of the rope diminishes the tension transmitted to the next part. It is found that after one turn the tension is diminished to about $\frac{1}{6}$, after two turns to $\frac{1}{6}$ of $\frac{1}{6}$ and so on. At this rate after five turns a pull of one pound weight on the free end would counteract a force of 4 tons at the other end.

The laws of friction were first investigated by Coulomb and are sometimes called by his name.

127. Slip on an Incline. When a body rests in an inclined plane the tilt of which is gradually increased there is some angle i at which slipping begins.

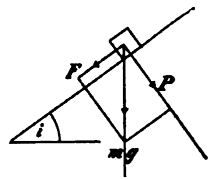


FIG. 66.

The weight of the body is mg and acts vertically. It may be resolved into a component $mg \sin i$ down the plane and a component $mg \cos i$ perpendicular to the plane. The latter component causes pressure between the surfaces, while the former is the force parallel to the surface which produces motion. Hence from the definition of this coefficient of static friction

$$\mu = \frac{F}{P} = \frac{mg \sin i}{mg \cos i} = \tan i$$

Thus the coefficient of static friction is equal to the tangent of the angle of slip. (This angle is also sometimes called the angle of repose.) This relation provides a simple method of measuring μ.

128. Slip on the Horizontal Produced by an Inclined Force. When a vertical force R is applied to the upper surface of a light body (such as a chip) which rests on a horizontal surface, slipping does not take place, but, when the force is inclined away from the vertical through a certain angle, the body slips. To find the magnitude of i, resolve R into a vertical component $R \cos i$, and a horizontal component $R \sin i$.

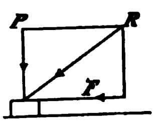


FIG. 67.

The former produces a pressure, P, between the surfaces, the latter is parallel to the surfaces and is the force, F, that causes motion. Hence

$$\mu = \frac{F}{P} = \frac{R \sin i}{R \cos i} = \tan i$$

It may be noticed that the cases of slipping treated in this section and the two preceding are really the same, as is shown by the similarity of the figures. In each the resultant force makes with the normal to the surfaces an angle i which, when slipping just begins, is such that $\mu = \tan i$.

129. Kinetic Friction. To keep one body sliding on another at a

constant speed a certain force, F, parallel to the surface of contact is required. Through a considerable range of speed this force is practically constant. The opposing resistance offered by the surfaces is called *kinetic friction.* It is found to be, for a certain pair of surfaces moving in a definite direction, proportional to the pressure, P, between the surfaces. Denoting the coefficient of proportionality by μ' we have

$$F = \mu' P$$

The constant μ' is called the *coefficient of kinetic friction.* It may be defined as *the ratio of the kinetic friction between two surfaces to the pressure between them.*

As in the case of static friction, for a given pressure between two surfaces the kinetic friction is independent of the area of contact.

The kinetic friction between two surfaces is in general less than the maximum static friction. The reason probably is that time is not allowed for the surface to settle into as close contact as if they were at rest. Moreover, kinetic friction is not quite independent of velocity. When the velocity is decreased until it is very small (how small depends upon the particular surfaces) the friction increases and it continues to increase as the velocity diminishes toward zero, and at a sufficiently small velocity the kinetic friction probably does not differ appreciably from the maximum static friction. At very great velocities the friction is generally less than at moderate velocities.

A friction dynamometer is a machine for measuring the power of an engine; the engine drives a wheel over which a belt hangs under known tension. From the tension of the belt and the number of revolutions made by the latter the work done is calculated.

When a lubricant is used between two surfaces there is no longer friction of solid on solid and the laws of kinetic friction no longer hold; the coefficient of friction depends on both pressure and velocity and the action is very complex. The friction of a skate on ice is probably greatly diminished by the momentary liquefaction of the ice immediately under the skate due to the great pressure exerted by the latter on a small area (see the part of this work on "Heat").

130. Sliding on an Inclined Plane. A body sliding down an inclined plane (Fig. 21) is urged downward by the component of its weight along the plane and retarded by friction. If the in-

clination of the plane to the horizontal is i, the component of gravity along the plane is $mg \sin i$. The pressure perpendicular to the plane is $mg \cos i$; hence the force of friction is $\mu' mg \cos i$. If the component of gravity down the plane exceeds the force of friction, the body will slide with an acceleration a. Hence, taking the direction down along the plane as the positive direction, we have by Newton's Second Law

$$ma = mg \sin i - \mu' mg \cos i$$

or

$$a = g \sin i (1 - \mu' \cot i)$$

This relation suggests a method of finding μ' by measuring a and i.

131. Rolling Friction. The term friction is also applied to the resistance experienced by a wheel in rolling on a surface without any slipping. The cause of the resistance is in this case entirely different. This is seen by considering the rolling of a heavy wheel on a soft substance, such as India rubber. If the wheel were at rest it would sink into the rubber, raising a small mound on each side of the contact. When the wheel is moving forward the mound is chiefly on the forward side at A. The pressure, P, of the rubber on the wheel at A is inclined to the vertical, in some such direction as AP. The point about which the wheel is momentarily rotating (§ 68) is not C but B in the figure, and the moment of P about B is necessarily opposed to that of F.

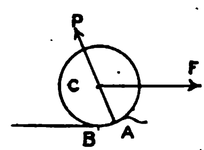


Fig. 68. Resistance to rolling.

It will be noted by the explanation that the resistance to the motion is greater the softer the surface, greater the greater the pressure of the wheel on the surface and less the larger the wheel, since a larger wheel will distribute its pressure over a larger surface and will not sink so deeply. When the surface on which the wheel rolls is hard very little deformation will ensue, and the resistance to the motion will be much less. Thus the resistance to the rolling of iron on india rubber is about ten times greater than the rolling of iron on iron. With a lignum vitæ cylinder of 16-in. diameter loaded with 1000 lbs. the rolling friction has been found to be about 3 per cent. of the sliding friction when the wheel was not allowed to rotate. Because of this difference rolling is, when possible, preferred to sliding. Thus rollers beneath a heavy body and the balls in a ball-bearing greatly diminish frictional resistance.

A pneumatic tire on a bicycle or automobile flattens out in contact with the ground and does not sink in, so that it gives the wheel the advantage of a much larger wheel. But it also bulges a little in front of the flattened part and this bulge is an obstruction of the same nature as the little mound in Fig. 68. On a perfectly smooth, plane, hard road a pneumatic tire would be a disadvantage. On a soft rough road it is a great advantage. For a hard smooth road the tire should be pumped "hard" for a soft road it should be "soft."

SIMPLE MACHINES.

132. Machines. A machine is a contrivance for applying energy to do work in the way most suitable for a certain purpose. The machine does not create energy; no machine can do that. To do work it must receive energy from some store of energy, and the greatest amount of useful work it can do cannot exceed the energy it receives.

Different machines receive energy in different forms, some in the form of mechanical (kinetic and potential) energy, some in the form of heat energy, some in the form of chemical energy, and so on. We shall only consider here machines which employ mechanical energy and do work against mechanical forces.

In certain very elementary machines, the so-called *simple machines,* the agent which supplies the energy exerts but a single force and the machine, at least as regards the useful work which it performs, is opposed by a single resisting force. The former is frequently called the "power"; but, to avoid confusion, we shall call it the *applied force.* The resisting force is frequently called the "weight"; but, as the opposing force is not always that of gravity, we shall call it the *resistance.*

Every machine in its action encounters a certain amount of frictional resistance; the work done against it is not usually useful work. This in many cases is very small, and, in treating (to a first approximation) of the simple machines, it is customary to neglect it.

133. Mechanical Advantage. The work done by the applied force, P, is measured by the product of P and the distance, p, through which P acts. The work done against the resistance is measured by the product of the resistance, Q, and the distance, q, through which it is overcome. In a simple machine (where fric-

tion may be neglected) these must be equal. Hence

$$\frac{Q}{P} = \frac{p}{q}$$

Hence p is as much greater than q as Q is greater than P. This principle was first stated by Stevinus. It is frequently put in the form "*what is gained in power* (*i. e. force*) *is lost in speed.*"

The ratio of Q to P for a machine is called the ***mechanical advantage*** of the machine. Since, for a perfect machine, that is one in which friction is negligible, the above ratio is also the ratio of p to q, it follows that we can deduce the mechanical advantage of such a machine from the ratio of the speeds without considering the inner mechanism of the machine.

134. Efficiency. By the *efficiency* of a machine is meant the ratio of the useful work, or work of the kind desired, to the energy received. For a simple machine without friction this would be unity. When there is friction the efficiency may have any value less than unity.

135. Levers. A lever is a bar supported at a point called the *fulcrum*, *F*; a force, *P*, applied to the bar at a point *A* will overcome a resistance, *Q*, acting at another point *B*. We shall suppose that *P* and *Q* act at right angles to the bar and to the axis of rotation at the fulcrum.

To find the relation between *P* and *Q* suppose the bar to turn through a very small angle, so that *A* moves through a distance

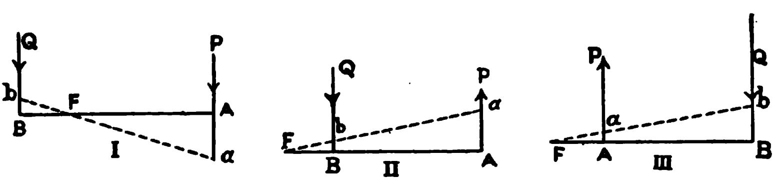


FIG. 69. Three classes of levers.

Aa and *B* through a distance *Bb*. The work done by *P* is $P \cdot Aa$ and the work done against *Q* is $Q \cdot Bb$. The conservation of energy requires that these should be equal. Hence

$$\frac{Q}{P} = \frac{Aa}{Bb} = \frac{AF}{BF}$$

This relation may also be found by considering the parallel forces acting on the bar or by taking moments about the fulcrum.

Levers are usually divided into three classes represented by the figures. In levers of the *first class* the force, P, and the resistance, Q, are on opposite sides of the fulcrum, and the resistance may be greater or less than the applied force. To this class belong the crow-bar, forceps, scissors, poker, and the common balance.

In levers of the *second class* the applied force and the resistance are on the same side of the fulcrum, the former being farther from it than the latter. Thus the resistance is always greater than the applied force. This class includes the oar of a boat, a pair of nut-crackers, a claw-hammer for extracting nails, etc.

In levers of the *third class* the applied force and the resistance are on the same side of the fulcrum, the former being nearer to the fulcrum than the latter. The purpose of such a lever is a gain of displacement or of speed. This class includes the forearm which is hinged at the elbow and acted on by the biceps at a distance of two or three inches from the elbow, a pair of tongs and the lever of a safety-valve for steam pressure.

136. The Wheel and Axle. A straight lever cannot raise a weight higher above the fulcrum than the distance of the weight from the fulcrum. The apparatus called a "wheel and axle" acts on the same principle as a lever but its range is not so limited. It consists of a wheel of large radius rigidly connected to an axle of smaller radius. The applied force, P, acts on a cord wrapped around the wheel, while the weight or resistance acts on a cord wrapped around the axle. The principle involved is that of a lever of the first class, the radius, R, of the wheel being the lever arm for the applied force, while the radius, r, of the axle is the lever arm of the resistance. Hence

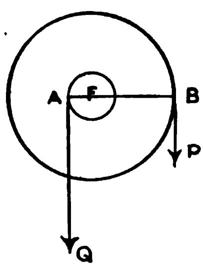


FIG. 70. Wheel and axle.

$$\frac{Q}{P} = \frac{R}{r}$$

This formula may also be proved directly by equating the work done by P in one complete revolution, $2\pi RP$, to the work done against Q, $2\pi rQ$; also by taking moments about F.

The principle of the Wheel and Axle is applied in the pilot wheel and in the capstan where the wheel is replaced by spokes in the axle, and in the winch, where there is but a single spoke, the crank arm.

In the above we have neglected friction, which is always considerable.

137. Differential Wheel and Axle. To obtain a very high mechanical advantage the wheel would have to be made very large, which would be inconvenient, or the axle would have to be made very small, which would greatly weaken it. To avoid these disadvantages the axle is made in two parts of different size and the cord is wrapped in the same direction around both, as indicated in the figure, the weight being carried by a pulley through which the cord passes.

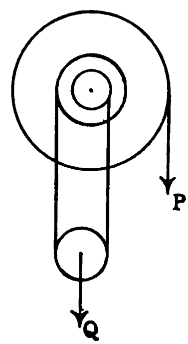


FIG. 71. Differential wheel and axle.

Let the radius of the wheel be R, that of the large part of the axle r, and of the small part r'. The upward force on the pulley is twice the tension of the cord and the downward force is Q, the weight of the pulley being neglected. Hence, by the principle of forces in equilibrium, the tension in the cord is $\frac{1}{2}Q$. In one revolution P does work $P \cdot 2\pi R$ and the tension of the cord acting on the smaller part of the axle does work $\frac{1}{2}Q \cdot 2\pi r'$, while work $\frac{1}{2}Q \cdot 2\pi r$ is done against the tension in the cord acting on the larger part of the axle. Hence

$$P \cdot 2\pi R + \tfrac{1}{2}Q 2\pi r' = \tfrac{1}{2}Q \cdot 2\pi r$$

$$\therefore \frac{Q}{P} = \frac{2R}{r - r'}$$

138. Pulleys. The simplest pulley is a wheel for the purpose of changing the direction in which a force is applied. It consists of a wheel in a framework or block which is either fixed or free. If it is fixed, the direction of the force is changed without any change in the magnitude (see Fig. 72).

If it is free and the two parts of the cord are parallel, the tension in any part of the cord is (neglecting friction and the weight

of the cord) equal to the force applied at its free end. Hence for equilibrium

$$Q = 2P$$

If the weight of the pulley is not negligible it may be included in Q. This formula is also readily found by the principle of energy; for each unit of length that Q moves P must move two.

139. Block and Tackle. Several pulleys are frequently used in combination so as to secure higher mechanical advantages. The most common arrangement is called the *block and tackle*. The

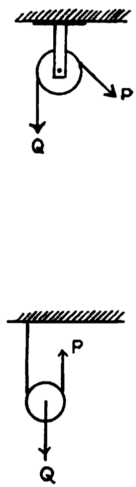


Fig. 72.

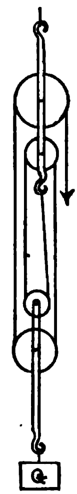


Fig. 73. Block and tackle.

pulleys are in two blocks with several pulleys in each block. The fixed end of the cord may be attached to either the upper or the lower block; if to the former, there will be an equal number of pulleys in the two blocks, as in the figure; if to the latter, there will be one more pulley in the upper block. When the distance between the blocks is decreased by one unit of length, each branch of the cord in contact with the lower pulley must shorten one unit of length. Hence

$$Q = nP$$

where n is the number of branches of the cord at the lower block.

140. The Differential Pulley or Chain Hoist. In this the upper block holds two pulleys of different diameters fixed rigidly to the same axis, while the lower block is replaced by a single pulley. An endless chain passes over the three pulleys as shown in the figure and is prevented from slipping by teeth on the pulleys. This is essentially a modification of the differential wheel and axle in which the wheel and the larger part of the axle have the same radius. The relation between P and Q, which may be worked out independently or may be obtained by putting $R = r$ in the formula of § 137, is

$$\frac{Q}{P} = \frac{2r}{r - r'}$$

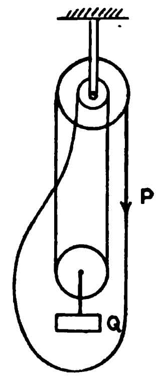


FIG. 74.

141. The Inclined Plane. A force less than the weight of a body may suffice to draw the body up an inclined plane. Let P be the force and W the weight (Fig. 75*a*). Also let h be the height and l the length of the plane. When the body has been

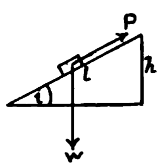


FIG. 75 a.

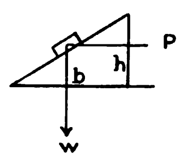


FIG. 75 b.

drawn up the whole length of the plane the work done by P (neglecting friction) will be Pl and the work done against W or the increase of potential energy will be Wh. These must be equal. Hence

$$\frac{W}{P} = \frac{l}{h}$$

This is essentially the same expression as already found (§ 50) by considering the component of W down the plane. If friction cannot be neglected the work done against it will be Fl, where F is the force of friction, and in the above equation P must be replaced by $(P - F)$.

If P act horizontally (Fig. 75b) the work done by P will be Pb. Hence (neglecting friction)

$$\frac{W}{P} = \frac{b}{h}$$

142. The Screw. The thread of an ordinary screw makes a constant angle with the length of the screw. If the thread of a vertical screw were supposed unwrapped, with its inclination kept constant, it would be an inclined line. The pitch of a screw is the distance, parallel to the length of the screw, between consecutive turns of the thread. The pitch divided by the outer circumference is the tangent of the inclination of the thread to the length of the screw.

If a nut carrying a heavy weight be turned around a vertical screw so that it ascends, the process will be similar to forcing a heavy body up an inclined plane by a horizontal force. In the jackscrew for raising heavy bodies the nut is fixed while the screw is turned by a lever. The useful work performed by the screw in one turn is the product of the resistance it overcomes, Q, and the rise in one turn, which is the pitch h. The work done in the same time is the product of the applied force, P, and the circumference, $2\pi R$, of the end of the lever arm. Equating these would give us a relation between P and Q; but friction is in general so large as to render the relation inapplicable.

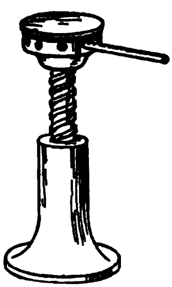


FIG. 76. Jackscrew.

GRAVITATION.

143. Law of Universal Gravitation. Until the time of Newton the weight of a body, or the measure of its tendency to fall to the earth, was generally regarded as an inherent property of matter that needed no further explanation. To Newton (and to some of his contemporaries) it occurred that the weight of a body on the surface of the earth is due to a force of attraction between the body and the earth, and that this attraction is only a particular case of a universal attraction between all bodies no matter where situated. Newton then sought to discover the law that such a

force would have to follow to account for the facts, how it would have to depend on the masses of the bodies and their distance apart. Now it was not possible for him to change the distance between a body and the center of the earth by any except an exceedingly small proportion, and the force between two bodies of ordinary size on the surface of the earth was so small that it escaped detection until a much later date. Hence he turned his attention to the motion of the moon and the planets.

Before the time of Newton, Kepler had, by a very extensive and painstaking study of the motions of the planets, arrived at certain laws known as Kepler's Laws. These may be stated as follows:

1. The areas swept over by a line joining a planet to the sun are proportional to the times.

2. Each planet moves in an ellipse in one focus of which the sun is situated.

3. The squares of the periods of revolution of the planets are proportional to the cubes of the major axes of the ellipses.

From these laws Newton showed that the motions of the planets could be accounted for on the supposition that between each planet and the sun there is a force of attraction, proportional to the product of the masses and inversely as the squares of their distances apart.

Newton also showed that, if we suppose that there is a force according to this law between every two particles, a sphere that is either homogeneous or may be regarded as made up of shells each of which is homogeneous will attract an outside body as if the sphere were concentrated at its center. The earth is very nearly such a sphere and must, therefore, according to the law of gravitation, attract (approximately) as if concentrated at its center.

144. Motion of the Moon. As evidence for the law of gravitation Newton showed that it correctly accounts for the motion of the moon. At the surface of the earth a body is attracted by the earth as if the latter were concentrated at its center. Now the radius of the earth is approximately 4,000 miles and the average value of the acceleration of a falling body may be taken as 32.2 feet per sec^2. The distance of the moon from the earth, which is somewhat variable, may be taken as approximately 240,000 miles or 60 times the radius of the earth. Hence, according to the law

of gravitation, the acceleration of a body at the distance of the moon due to the earth's attraction should be $32.2/60^2$ or .00894 ft. per sec^2.

The acceleration A of the moon towards the earth (§ 32) equals v^2/R. The period of rotation of the moon, also slightly variable, is about 27 days, 8 hours. Calling this T, we have $v=(2\pi R/T)$. Hence $a=(4\pi^2R)/T^2$, or reducing R to feet and T to seconds $a=.00896$. This value of A, calculated from the observed period of the moon, agrees as closely with the preceding value, deduced from the law of gravitation, as could be expected when the fact is considered that only approximate values for the various constants have been used. The argument must be considered very strong evidence for the law of gravitation.

145. Force of Gravitation Proportional to Mass. According to the law of gravitation the attraction between two bodies is proportional to their masses and is independent of the materials of which they consist. One proof of this was given by Galileo, when he dropped two cannon balls of different sizes from the leaning tower of Pisa and found that they reached the ground in very nearly the same time. Their accelerations being equal, the ratio of the force to the mass, must, according to the second law of motion, be the same for both. Yet in Galileo's experiment the larger weight was slightly ahead of the smaller, and Galileo correctly explained this difference by remarking that the air-friction would be proportionately less on the larger body. In fact, because of this air friction and the rapidity of the motion, it would be difficult to give a very convincing proof of the law by means of bodies falling with the full acceleration due to gravity.

To avoid this difficulty Newton experimented with a pendulum, the motion of which depends on gravity but on a fraction only of the full force of gravity, namely, the component along the arc of vibration. The bob of the pendulum was a thin shell and into this he put in successive experiments different substances. In each case the same weight, as tested by weighing with a balance, was put into the box and, since the force of air-friction on the box for the same amplitude of vibration would be the same no matter what the contents of the box, it followed that at a given inclination to the vertical the force causing the motion would be

always the same. He found that the time of vibration was always the same no matter what the contents of the box and hence the masses must also have been the same; that is, equal masses of different substances have equal weights. These experiments were afterward repeated by Bessel with much greater care and with the same result.

The above experiments prove that gravitation is not, like magnetic attraction, a force that depends on some quality of a body other than its mass, that is, not a selective force but a general force. That it does not depend on any other physical condition such as temperature, or on any chemical condition such as molecular combination, has also been shown by most careful weighing. A third body placed between two bodies has not the least effect in shielding them from their mutual attraction. The fact that a lump of gold, when hammered out into an exceedingly thin sheet, suffers no change of weight shows that the weight of a body does not depend on its form, that gravity acts on the particles whether surrounded by other particles of the same kind or not.

146. The Constant of Gravitation. The law of gravitation may be stated as a formula, viz.

$$F = G\frac{mm'}{r^2}$$

where G is a constant number called *the constant of gravitation.* To find the magnitude of G it is necessary to measure F in some case where m, m' and r are all known. This was first done by Henry Cavendish in 1797–8, and the experiment, usually called the *Cavendish experiment,* has been repeated many times since with increasing care and accuracy. Cavendish suspended two balls, A and B, from the ends of a long light horizontal rod which was supported by a long fine wire attached to the middle, C, of the rod. On opposite sides of this and at known equal distances he placed two large spheres of lead, P and Q. The attraction between each ball and the adjacent large sphere had a moment about C that produced a twist of the supporting wire. When the spheres

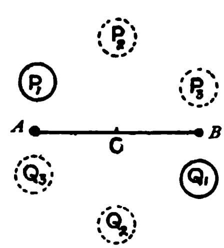


Fig. 77. Principle of the Cavendish experiment.

were in the position P_1Q_1, the twist was in one direction and when they were in the position P_2Q_2 the twist was in the opposite direction. To deduce the force of attraction from the magnitude of the twist, the constant of torsion of the wire (§ 119) had to be found by timing vibrations of the wire, when the spheres were in the position P_2Q_2 where they had no influence on the vibration of AB. Thus F, m, m', and r were found and when they were substituted in the above formula the value of G was obtained.

In more recent work the apparatus has been greatly improved. The greatest improvement has been in the substitution of very fine quartz thread for the wire. This also permitted of the apparatus being greatly reduced in size, so that, whereas AB in Cavendish's experiment was 6 ft. long, in Boys' apparatus it was only .9 inch, and the masses A, B, and P, Q, were also greatly reduced in size. The value obtained for G (using C.G.S. units) was 6.6579×10^{-8}; this is, therefore, the force in dynes of the attraction between spheres of one gram each at a distance of 1 cm. between their centers. When it is remembered that a dyne is about the weight of a milligram, it is seen that the force measured in the above experiments must be exceedingly small; hence the difficulty of the experiment.

147. The Mean Density of the Earth. The determination of G made it possible to calculate the mass of the earth (hence Cavendish is sometimes said to have been the first to "weigh the earth"). For if m' in the formula for the law of gravitation be put equal to one gram and m and r be taken as respectively the mass and radius of the earth, F will be the force of attraction between the earth and a body of 1 gm. mass and this is, as we know, 980 dynes. Thus the formula gives us the value of m, the mass of the earth, which is found to be 5.97×10^{27} gms. This figure is so large as to convey no distinct meaning, but a different way of stating the result will be more easily comprehended. The density of a homogeneous body such as water is its mass per unit volume, and when the density of a body is not everywhere the same we may speak of its *mean density* or its whole mass divided by its whole volume. Thus to get the mean density of the earth we divide its whole mass, as given above, by its whole volume. The result is 5.527, that is to say, on the average the earth is 5.527 times as dense as water. It is remarkable that Newton, on what evidence we cannot now conjecture, supposed the mean density of the earth to be between 5 and 6.

148. The Tides. At any place on the shore of an ocean the level of the water rises to a maximum and falls to a minimum once in about every twelve hours and 25 minutes. These risings and fallings are called the *tides*. They are due to the forces of attraction which the moon and the sun exercise on the water on the surface of the earth and to the rotation of the earth. The complete explanation of their action is extremely difficult, owing to the irregularities of the continents and to other causes.

UNITS.

149. Fundamental and Derived Units. The measurement of any quantity consists in comparing it with a unit of the same kind (§ 2). Thus a length is measured by comparing it with a unit of length, such as the foot or meter; a velocity is measured by comparing it with a unit of velocity, such as a foot per second and so on. Hence we need as many units as there are different kinds of quantities to be measured.

But all these necessary units are not necessarily independent. It is found that in Mechanics three independent or *fundamental* units are sufficient; all others can be defined in terms of these. A unit defined by reference to some other unit or units is called a *derived unit.*

150. Absolute Systems of Units. A system of units in which the derived units bear the simplest possible relation to the fundamental units is called an *absolute* system. In such a system the unit of area or surface is the square of the unit of length, the unit of volume is the cube of the unit of length, the unit of velocity is a velocity of unit length per unit time, and so on. Given any three fundamental units of length, time and mass, we can build up an absolute system of derived units. Thus we have one absolute system founded on the cm., gm., and sec., another founded on the ft., lb., and sec., and so on.

151. Dimensions of Units. It is sometimes necessary to translate results from one absolute system to another. It then becomes necessary to consider how the magnitude of a derived unit changes when the fundamental units are changed. For this purpose we need to know the dimensions of the derived unit, that is, the powers of the fundamental units to which the derived unit is proportional. For instance the unit of area is

the square of the unit of length, or area is of 2 dimensions in length, a statement briefly summarized by the *dimensional formula* $[A] = [L]^2$; similarly using [Vol] for the unit of volume $[\text{Vol}] = [L]^3$.

152. Dimensions of Velocity. The unit of velocity is defined in terms of the unit of length and the unit of time. To find the dimensions in these units consider any relation between velocity, length, and time, such as $s = vt$ (§ 19). This is a relation between numerical measures (§ 2), but it implies certain relations between the units used in measuring these quantities; both sides must be of the same dimensions in fundamental units, or they could not be equal. Hence if we denote the unit of velocity by $[V]$, $[L] = [V][T]$ or $[V] = [L][T]^{-1}$. Thus velocity is of $+1$ dimension in length and -1 dimension in time.

153. Dimensions of Acceleration. Consider any relation between acceleration length and time, such as $s = \frac{1}{2}at^2$. From this by the line of reasoning explained in the last section we derive at once $[L] = [A][T]^2$. Hence $[A] = [L][T]^{-2}$. The sign of equality in such expressions denotes equality of dimensions. Constant numerical factors (such as the $\frac{1}{2}$ above) are of zero dimensions, that is, they do not change when we change the fundamental units.

154. Other Derived Units. The above examples sufficiently explain the method by which the following table is derived.

TABLE OF DERIVED UNITS USED IN MECHANICS.

Quantity.	Relation of Numerics.	Dimensions of Units.	Name of Unit in C.G.S. System.
Linear velocity, v	$s = vt$	$[L][T]^{-1}$	
Linear acceleration, a	$s = \frac{1}{2}at^2$	$[L][T]^{-2}$	
Angular velocity, ω	$\phi = \omega t$	$[T]^{-1}$	
Angular acceleration, α	$\phi = \frac{1}{2}\alpha t^2$	$[T]^{-2}$	
Force, F	$F = ma$	$[L][T]^{-2}[M]$	dyne
Moment of Force, L	$L = Fp$	$[L]^2[T]^{-2}[M]$	
Moment of Inertia, I	$I = mr^2$	$[L]^2[M]$	
Work, W	$W = Fs$	$[L]^2[T]^{-2}[M]$	erg
Kinetic Energy, E.	$K.E. = \frac{1}{2}mv^2$	$[L]^2[T]^{-2}[M]$	erg
Potential Energy	$P.E. = Fs$	$[L]^2[T]^{-2}[M]$	erg

155. Examples of Use of Dimensional Relations. Where a derived unit has no particular name its dimensional formula is a sufficient name. Thus the unit of acceleration has no special name and 10 units of acceleration in the C. G. S. system is written

$$10 \frac{\text{cm.}}{\text{sec.}^2}$$

A frequent use of dimensional relations is in changing the measure of quantity from one absolute system of units to another. For example, the acceleration of gravity is $980 \frac{\text{cm.}}{\text{sec.}^2}$, what is it in $\frac{\text{meter}}{\text{min.}^2}$? Suppose it is x.

Then

$$x\,\frac{\text{meter}}{\text{min.}^2}=980\,\frac{\text{cm.}}{\text{sec.}^2}\;\therefore\;x=980\,\frac{\text{cm.}}{\text{m.}}\cdot\left(\frac{\text{min.}}{\text{sec.}}\right)^2=35{,}280.$$

Another use of these relations is in testing the accuracy of complicated formulas. The two sides of the equation must be of the same dimensions or they could not stand for the same thing.

PROPERTIES OF MATTER.

Constitution of Matter.

156. In the preceding chapters on the principles of Mechanics, we have had (with slight exceptions) to consider matter from but one point of view, namely, its inertia. The forces that the particles of a body exert on one another did not need to be considered, for they cancelled out when the action of the body as a whole was considered.

We shall now consider other important properties of matter, especially those which depend on the forces between particles. It will be seen that the connections between these properties are not so well understood as the relations between the quantities studied in Mechanics. This is chiefly because the ultimate particles of a body are so small that they cannot be studied separately. In fact we can only infer their existence and relations from the properties they exhibit in large groups.

157. The Three States of Matter. Following popular language we classify bodies as solids and fluids. The characteristic of a *solid* is that it has a definite shape which it does not readily relinquish, while a *fluid* flows easily or changes its shape in response to the smallest influence. (It will be seen later that the distinction is not quite definite, that some bodies lie on the borderland between the two classes.) The particles of a solid are held in (practically) fixed positions by the forces between them, but each particle has a freedom to vibrate about its mean position (see § 161).

Fluids are divided into *liquids* and *gases*. The peculiarity of a liquid is that, while it readily flows, it has a definite volume which it does not readily change. A gas yields to the smallest force exerted to change its volume, in other words, it has no definite volume of its own, but takes the volume of the containing vessel however large. (This distinction also is only general.) The particles of a liquid are close together and attract each other with powerful

forces. These forces react strongly against outside forces that tend to change the mean distance between the particles, but they are such as to permit sliding motions of the particles. The particles of a gas are practically separate bodies flying in space and exerting no appreciable forces on one another except at impact of particle on particle.

158. Elements and Compounds. In innumerable cases two or more substances coalesce to form a new substance that may be so distinct in all its properties that nothing apparently remains to suggest the constituents from which it was formed. Thus two substances, oxygen and hydrogen, gaseous under ordinary conditions, combine to form a liquid, water. Harmless substances may on combination form deadly poisons or explosives. Substances that may be made from constituents which have properties distinct from the resultant are called *compounds*.

Conversely, compounds may be divided up into constituents differing widely from the original substance and these constituents may be themselves capable of being resolved into other constituents. But there are many substances which have not as yet been resolved into constituents and such are called *elements*. Of these there are about 80 known.

159. Molecules and Atoms. Many facts, chiefly such as are more closely studied in chemistry, justify the belief that (1) an element consists of very small particles called *atoms*, (2) all the atoms in one elementary body are identical in size and other properties, but different from those of any other elementary body, (3) these atoms are combined in similar groups called *molecules* (in some substances the atom and the molecule are identical). It is also believed that a compound consists of molecules and that each molecule consists of two or more atoms of the constituents of the compound. There is also much reason to believe that in many substances, especially liquids and solids, molecules are frequently combined to form groups or molecular aggregates of two or more molecules each.

Molecules and atoms are extremely small and will probably never be separately visible, however much optical instruments may be improved. Thus in a cubic centimeter of a gas under ordinary conditions there are about 4×10^{19} molecules.

160. Intermolecular Forces. It is evident from the great forces necessary to pull a solid body apart that there are comparatively great forces between particles; but the ease with which a brittle body falls apart when a slight crack appears shows that the forces are only appreciable when the attracting particles are very close together. The latter point is also shown by the fact that a body molecular forces are, of course, different for different substances, and the characteristic properties of different substances probably depend on these differences.

Roughly speaking, it may be said that the force of molecular attraction is inappreciable at distances greater than about $\frac{1}{2000000}$ cm. or $\frac{1}{5000000}$ inch. The magnitude and the range of the intermolecular forces are, of course, different for different substances, and the characteristic properties of different substances probably depend on these differences.

161. Kinetic Theory of Matter. There is very strong evidence that the particles of which bodies are made up are in no case at rest. Thus two different gases contained in two different vessels mix with great rapidity when the vessels are put in communication. This process is called *diffusion*. Liquids will also diffuse into one another (except non-miscible liquids like oil and water), though much more slowly than gases, because of the greater closeness of the particles and the frequent changes of direction of vibration of a particle produced by impact on other particles. Even many solids show by diffusion that their particles are not at rest; thus when a small block of gold is placed on a block of lead with planed surfaces in close contact, after the lapse of some weeks it is possible to detect particles of gold which have wandered into the lead and vice versa. There are many other reasons for believing that the particles of matter are in all cases in motion. This hypothesis is called the hypothesis of the *kinetic constitution of matter*.

162. Density and Specific Gravity. The *density* of a body is its *mass per unit volume*. If the masses of all equal volumes of the body are the same, the density is *uniform* and equal to the mass in any unit of volume. If the masses of equal volumes are not the same, the density is not uniform. The *mean density* in any particular volume of the body is the mass in that volume divided by

the volume. The *density at a point* is the mean density in a small volume enclosing the point when the volume is supposed to be decreased without limit.

The measure of the density of a body depends of course on the units of mass and volume employed. If the C. G. S. system be employed density is the number of gms. per c.c. In this system the density of water at 4° C. is very nearly unity, since the gram was originally intended to be the mass of 1 c.c. of water at 4° C. In the British system the density of a body is the number of lbs. per cu. ft. of a body. In this system the density of water is 62.4, since that is the number of lbs. in a cu. ft. of water.

The *specific gravity* of a body is the ratio of its density to that of some standard substance. The standard usually employed is water at 4° C. Thus if D be the density of a body and d that of water at 4° C. the specific gravity of the body is D/d. Now in the C. G. S. system d is very nearly unity. Hence in this system density and specific gravity are practically identical. But in the British system, since d is 62.4, the specific gravity of a body is its density divided by 62.4.

Table of Densities.

Aluminium	2.60	Iron (about)	7.60
Brass (about)	8.50	Lead	7.86
Copper	8.92	Platinum	11.30
Gold	10.30	Silver	10.53

PROPERTIES OF SOLIDS.

163. Homogeneity and Isotropy. A **homogeneous** body is one which has at all points the same properties, so that small spheres of equal radii cut out of different parts of the body would be identical in properties. Many crystals are nearly perfectly homogeneous, and so, too, is good glass, such as plate glass or the glass of lenses. Many other bodies are approximately homogeneous, such for example are most metals, wood, stones, etc.

An **isotropic** body is one which has at any point the same properties in all directions, so that if at any point a sphere were cut out there would be nothing in the properties of the sphere to indicate the original direction of any diameter. All liquids and

gases are isotropic under ordinary conditions but many substances, such as crystals, woods and drawn metals, are distinctly non-isotropic.

164. Elasticity. When the shape or volume of a solid is changed by the application of some force, there is in most cases a tendency to return to the original shape or volume when this force is removed. This tendency to recover from distortion is called *elasticity*. It is one of the most important properties of a solid, since the usefulness of many bodies such as springs, musical instruments, etc., depends on the extent to which they possess this property. It is, therefore, a property that has been very extensively studied.

165. Strain. *Any change of shape or of volume or of both is called a strain.* Thus the bending of a beam, the twisting of a rod, the compression of a liquid or a gas into a smaller volume are strains. The term strain is a geometrical one and its definition contains no reference to force or energy, although, as we shall see, force and energy are present when a body is in a state of strain.

A strain that consists in *a change of shape only* without any change of volume is called a *shear,* since this is the kind of distortion that a pair of shears produces. The strain of a moderately twisted wire or rod is a shear.

A strain that consists in *a change of volume* only without any change of shape has not received any special name, but we may for brevity call it *a volume-strain.* Such, for example, is the strain of a sphere of cork or of any isotropic body when placed in a fluid which is subjected to great pressure in a closed vessel.

While for simplicity we have first enumerated strains in which either volume or shape alone changes, strains which involve changes of both are more common. Thus the stretching of a wire, the compression of a pillar, the bending of a beam, etc., are strains of both volume and shape. A body is said to be homogeneously strained or the strain is described as *homogeneous* when the nature and magnitude of the strain is the same at all points in the body. Thus when a wire is stretched or a rod compressed and when a liquid or gas is subjected to pressure, the strain is homogeneous. But when a wire or rod is twisted the strain is greatest at the surface and least at the center, and, when a beam is

bent, there is a stretching on the convex side and a compression on the concave side and the strain is heterogeneous.

166. Stress. When a body is in a state of strain owing to the action of external forces on it, there are internal forces between contiguous parts of the body in addition to whatever internal forces there may have been before the strain occurred. If at a point a dividing plane be imagined, the part of the body on one side will act with a certain force on the part on the other side and the latter will react with an equal and opposite force. These two forces together, the action and the reaction, constitute a *stress*. In some cases as we shall see the stress is perpendicular to the imaginary dividing plane and in others parallel to it, but in any case *the magnitude of the stress is the force per unit area* of such an imaginary dividing plane.

The terms *homogeneous* and *heterogeneous* apply to stress just as to strain. In many cases, for example in the stretch of a wire by an attaching weight, the stress in a body is equal to the external force per unit area that acts on the body and produces the strain, and in such cases we may speak of this external force per unit of area as the stress. In other cases, as, for example, in the bending of a beam by a weight acting at some point, the stress does not bear a simple relation to the external force and we must take care to distinguish them.

167. The Measure of a Strain and of a Stress. A strain which consists in a change of volume only is measured by *the proportion in which the volume is changed*. If the strain is homogeneous the measure may be taken as the change in unit volume, or if a volume V becomes $(V+v)$ the measure of the strain is v/V. If the strain is not homogeneous the measure of the strain at any particular point is the value of v/V at the point, when V is taken as the volume of an indefinitely small portion surrounding the point. To produce this change of volume force must be applied to the surface of the solid in the form of either a pressure or a tension, and inside the body each part will press or pull on each neighboring part. The amount of this pressure or pull per unit area is the measure of the stress.

The measure of a *shear* will be most readily understood by considering the simplest way in which a shear may be produced. Consider, for example, a rectangular block of a firm jelly between

two boards to which it adheres. Let *PQRS* be one rectangular face and *PQ, RS* the edges of the boards. Apply to the boards equal and opposite forces parallel to them and to the face *PQRS*. The face *PQRS* is changed to the form P_1Q_1RS. Each section of the block parallel to the boards moves parallel to itself a distance proportional to its distance from *RS*. Each of the right angles of *PQRS* is changed by the same amount, say θ, and this change is the measure of the shear.

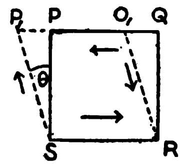


FIG. 78. Shear and shearing stress.

When θ is small, as it is in most practical cases, the magnitude of the angle θ in radian measurement is $P_1P \div PS$, or taking *PS* equal to unity, ***the relative displacement of two planes unit distance apart.***

If an imaginary plane be supposed drawn anywhere in the block parallel to the boards, the part on one side of this plane will exert a tangential force on the part on the other side and this force will equal the force applied to the boards. The ***magnitude of the force per unit area*** is the measure of the shearing stress.

While we have referred only to the forces parallel to *PQ* and *RS* it is clear that the shear cannot be produced without other forces applied to the block. If only the two forces described were applied, the block would not be at rest but in rotation since the two constitute a couple. The effect is readily perceived when the attempt is made to apply the two opposite forces. It is in fact necessary to also apply other forces forming an opposite counterbalancing couple, say along SP_1 and Q_1R. The effect of all four forces is to produce a stretch along RP_1 and a compression along Q_1S and the proportional stretch is equal to the proportional compression, since there is no change of volume.

168. Hooke's Law. When any body is strained beyond a certain amount and then released, it fails to return completely to its original form and volume or it retains a permanent set. The largest strain of any kind which a body may undergo and still completely recover from when released is called *the limit of elasticity* for that form of strain, and the corresponding stress is called the *limiting stress.* The limit of elasticity is, of course, widely different for different substances. Thus rubber may be greatly extended and yet recover, while the limit for glass and ivory is very small. (Cases in which the limit is somewhat indefinite will be considered later.)

Within the limit of elasticity a simple law, first stated by Hooke is 1676 and known as *Hooke's law,* holds, namely, "*stress is proportional to strain*" (Hooke's statement in Latin was "Ut tensio sic vis"). Hooke illustrated his law by various cases of strain, such as the stretching of a spiral spring, and of a wire, the bending of a beam, the twisting of a wire and so on.

169. Moduli of Elasticity. While elasticity has already been defined as the tendency of a body to recover its shape or volume when distorted, the definition is purely qualitative and affords no means of assigning a numerical value to the elasticity of a substance. A quantitative definition of the elasticity of a substance for any form of strain follows from Hooke's law. *The measure or modulus of elasticity is the ratio of the magnitude of the stress to that of the accompanying strain,* this ratio being a constant within the limits of elasticity. As there is a great variety of forms of strain there is a correspondingly large number of moduli of elasticity for any substance; but only a few of these are important enough to be enumerated.

When the strain is one of volume only the elasticity is called *elasticity of volume.* The modulus of elasticity of volume or the *bulk-modulus,* as it is frequently called, is the ratio of the stress, or the pressure per unit area, P, to the change of volume per unit volume. The bulk modulus of a substance is usually denoted by k. Hence, if a volume V undergoes a change of volume v and the stress is P,

$$k = \frac{P}{\frac{v}{V}} = \frac{PV}{v}$$

The reciprocal of the bulk modulus is called the coefficient of *compressibility* of the substance. It means the ratio of the proportional compression to the pressure per unit area, or supposing the latter to be unity, the coefficient of compressibility is the ratio in which the volume is reduced by unit pressure per unit area.

When the strain is a shear the modulus of elasticity is called the *shear modulus,* or often the *simple rigidity,* and is the ratio of the shearing stress to the shear. Denoting the shearing stress by T, the shear corresponding to T by θ, and the shear modulus

by n,

$$n = \frac{T}{\theta}$$

170. Torsion. When a wire or rod of homogeneous isotropic material is twisted, we may imagine the whole length divided into transverse slices of equal thickness by planes perpendicular to the axis. Each such slice will be rotated about the axis to an extent proportional to its distance from the fixed end. Moreover one face of each slice (the one farthest from the fixed end) will be rotated more than the other. Let us now suppose that each slice is very thin, and that it is divided up before twisting into very small cubes (or nearly cubes) by a series of imaginary planes through the axis intersected by concentric cylinders. Thus each cube will have four edges parallel to the axis, four others in the direction of radii while the remaining four will be short and practically straight arcs of circles. After the twist each cube will have a strain like the cube of jelly in § 167. Hence the strain is a shear but, since the strain of each cube will be proportional to its distance from the axis, the strain is not homogeneous.

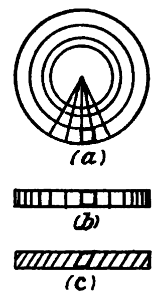


Fig. 79. Shear of a small cube in a twisted wire.

The *constant of torsion* of a wire has already been defined in § 119.

171. Young's Modulus. A very frequent form of strain is that of a uniform wire or rod which is clamped at one end and is acted on by a longitudinal force at the other end. Such a strain is called a *stretch*. Any short part of the wire is extended in the same proportion as the whole wire. The measure of the stretch is the extension per unit length or, denoting the unstretched length of the wire by L and the total extension by l, the stretch is l/L. The measure of the stress is the external pull per unit of cross-sectional area. Denoting by P the whole force applied to one end and by a the cross-sectional area of the wire, the pull per unit area anywhere in the rod due to the force P is P/a, which is, therefore, the measure of the stress. The value of Young's modulus, which we may denote by M, is, therefore, $(P/a) \div (l/L)$ or

$$M = \frac{PL}{al}$$

For some common materials the average values of k, n, and M in C.G.S. units are as follows:

	Dynes per cm^2. k.	n.	Lbs. wt. per sq. in. M.
Copper	17×10^{11}	4×10^{11}	11×10^{11}
Glass	4×10^{11}	2×10^{11}	6×10^{11}
Iron (wrought)	15×10^{11}	7×10^{11}	19×10^{11}
Lead	4×10^{11}	2×10^{10}	1×10^{11}
Steel	17×10^{11}	8×10^{11}	23×10^{11}

The reciprocal of M is sometimes called the coefficient of extensibility and is evidently equal to the proportion in which the length is increased by unit longitudinal force per unit area of cross-section.

172. Volume Changes when a Wire is Stretched. When a wire or rod is stretched there is obviously a change of shape in every part of the wire or rod, for the length is increased while the cross-section is decreased. Whether a change of volume also occurs can only be determined by experiment. If the cross-section diminishes in the same proportion as that in which the length increases, there is no change of volume; whereas if the proportion in which the length increases exceeds that in which the cross-section diminishes, there is an increase of volume. Careful experiment shows that in all cases there is an increase of volume; but in some substances, *e. g.*, india rubber, the change of volume is very small.

173. Flexure. A very common strain closely related to stretching is that of a plank supported at both ends and carrying a load at the middle, or supported at the middle and loaded at each end, or clamped horizontally at one end and loaded at the other end. A little consideration will make it clear that in these cases we have to do with stretches and shortenings such as those already discussed. If we suppose the plank divided into a large number of longitudinal strips, the strips on the convex side are stretched by the bending, while those on the concave side are shortened. There must of course be an intermediate surface where there is neither stretch nor compression and this surface is called the ***neutral surface.***

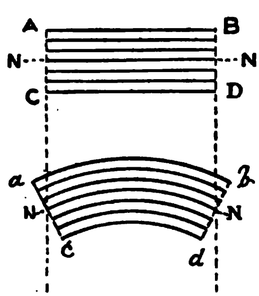


FIG. 80. Bending of a beam (exaggerated).

The extension or compression of any strip is proportional to its distance from the neutral surface. Thus the strain, while not homogeneous, is everywhere of the nature of an extension or a

compression and Young's modulus is the only modulus involved. If a bar of length l, breadth b and depth d be supported at both ends and be subjected to a perpendicular force F at the middle the depression produced is $Fl^3/4Mbd^3$.

174. Direct Impact of Elastic Bodies. When two bodies in motion collide each exerts a momentary force on the other and each, therefore, suffers a change of velocity. The result is difficult to calculate except in certain simple cases.

When the bodies are uniform spheres and are moving before impact along the line joining their centers, the result can be calculated. Let the masses be m and m' and the velocities before impact u and u', and suppose that both are in the positive direction and $u > u'$. After the impact m' will be moving faster than m. Let the velocity of m after impact be v and let that of m' be v'. Then $v' > v$. During the short time of contact each body exerts a force on the other and, by the Third Law of Motion, these forces are at any moment equal and opposite. These forces also act for the same length of time and must therefore produce equal and opposite changes of momentum. Hence the total momentum after impact equals the total momentum before impact, or

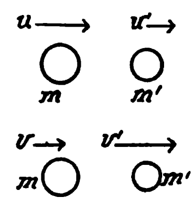


FIG. 81. Motion of spheres before and after impact.

$$mv + m'v' = mu + m'u' \tag{1}$$

If the problem be to find the velocities after impact this equation will not suffice, since it contains two unknown quantities v and v'. A second relation between v and v' was discovered experimentally by Newton. He found that for given materials the ratio of the speed of separation, $(v' - v)$, to the speed of approach, $(u - u')$, is a constant, which is (at least very nearly) independent of the masses and velocities of the bodies and depends only on their materials and the direction of the grain if they are not isotropic. This constant ratio is called the *coefficient of restitution*. Denoting it by e, we have

$$\frac{v' - v}{u - u'} = e$$

or

$$v - v' = -e(u - u') \tag{2}$$

From (1) and (2) v and v' can be calculated. For simplicity in establishing these equations we chose the case in which all the velocities are in the positive direction, but they are algebraic equations applicable to all cases. In applying them care must be taken to give the proper signs to the velocities.

Some simple deductions may readily be drawn. When e is zero, as it very nearly is for such soft substances as putty and lead, we see from (2) that v and v' are equal, or the bodies do not separate after impact.

If the masses of the spheres be equal and $e = 1$, the spheres will on impact exchange velocities. For in this case the two equations become

$$v + v' = u + u'$$
$$v - v' = -u + u'$$

Hence $v = u'$ and $v' = u$, which proves the statement.

If one of the bodies, say m', is of very great mass compared with the other and is initially at rest, its velocity after impact will be very small. Putting both u' and v' equal to zero in (2) we get

$$e = -\frac{v}{u}$$

This is the case when a small ball is dropped on a very large block. Let the height of fall be H and the height of rebound h. Then $u = \sqrt{2gH}$ downward and $v = \sqrt{2gh}$ upward. Hence

$$e = \sqrt{\frac{h}{H}}$$

This affords a simple experimental method for finding e.

175. Oblique Impact of Smooth Spheres. The impact of two spheres is described as oblique when the spheres are not moving before impact in the direction of the line through their centers. The lines of motion of the centers before impact may be in one plane, as when two equal balls rolling on a plane surface impinge, or these lines may be in different planes. In either case we may resolve the velocity of each ball before impact into two components, one in the direction of the line through the centers at the moment of impact, the other in a direction perpendicular to that line. Only the first component will be affected by the impact since (the spheres being

supposed frictionless) the only force will be a pressure in the line of the centers. The change in this component may be calculated as in the case of direct impact; then by compounding this component for each sphere after impact with the unchanged component we can find the motion of each sphere after impact.

When a smooth ball impinges obliquely with a velocity u on a fixed surface in a direction making an angle a with the normal, its component velocity parallel to the surface is $u \sin a$ and perpendicular to the surface $u \cos a$. If it rebounds with a velocity v in a direction making an angle b with the normal, the components become $v \sin b$ and $v \cos b$. The component parallel to the surface is not changed while that perpendicular to the surface is changed in the ratio $e : 1$. Hence

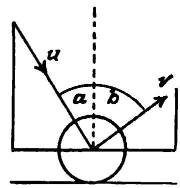


FIG. 82. Impact on a fixed surface.

$$v \sin b = u \sin a, \quad v \cos b = eu \cos a.$$

Hence by dividing corresponding sides

$$\tan b = \frac{\tan a}{e}$$

Thus the direction of rebound is more nearly parallel to the surface than that of impact. This is the basis of a method that has been employed for finding e.

176. Loss of Energy on Impact. The kinetic energy of two smooth spheres before impact and that after impact can be calculated from their masses and velocities. The total kinetic energy of two bodies is less after impact than before (except when e is unity) and other forms of energy, such as heat and sound, are produced.

177. Vibration of Elastic Bodies. When a body is strained within the limit of elasticity the internal stresses tend to restore the body to its original condition. When released from the external deforming force the body vibrates, and, since the restoring forces are at each stage proportional to the distortion, the vibrations are simple harmonic vibrations of a constant period. This, for instance, is the case when a rod firmly clamped at one end is bent and released. When the vibrations are sufficiently rapid, as is the case of the prongs of a steel tuning fork, sound is produced, and the ear can test the constancy of the period of vibra-

tion by the steadiness of the pitch; the vibrations gradually die down, that is, the extent of the maximum strain in each vibration decreases, yet the period remains unchanged, showing that within the limits of vibration the stress is, so far as the delicate sense of hearing can detect, accurately proportional to the strain. A tuning fork can be made of any metal, of wood or other solid substance; and, while the sound may in many cases be weak and short-lived, the steadiness of pitch while it lasts is an excellent proof of Hooke's law.

178. Strain Beyond the Elastic Limit. As an illustration of what happens when a substance is strained beyond the elastic limit, that is, beyond the range in which Hooke's law holds, we shall consider the stretching of a wire. When a force that stretches it beyond the limit is applied to it and this force is steadily increased, it elongates in greater proportion for each successive equal increase of the force. As the force is increased, at a certain strain, called the *yield point,* a very rapid increase of strain sets in at some point of the wire, and the strain at that point continues to increase, even if the force is not increased, until at last the specimen "necks in" and breaks. Beyond the yield point the substance flows much like a very viscous liquid. If during this process the force be diminished somewhat, the strain will still continue to increase but at a diminished rate; and, when the force is diminished sufficiently, the strain ceases to increase before breakage occurs. If at this stage the applied force be removed entirely, the wire will contract somewhat, but a large permanent set will remain. The wire will then act like a different wire with a new elastic limit.

179. Elastic After-effects. From strain within the elastic limit the strained material completely recovers in time and there is no permanent set; but frequently the immediate recovery on removal of the force is not complete, and there remains a small temporary set from which the material only slowly recovers. This slow recovery from temporary set is called an *elastic after-effect.* It is shown by rubber and glass and other substances which consist of mixtures of diverse molecules; but crystals and quartz threads do not show it.

It is readily demonstrated by clamping both ends of a rubber cord (used for tires of small wheels) and attaching a small mirror to the middle to reflect a beam of light on a scale. Such an arrangement will show a double after-effect due to successive twists in opposite directions.

180. Fatigue of Elasticity. The vibrations of a torsional pendulum are maintained by the elasticity of the wire; they slowly die away owing to air resistance and internal friction in the wire. If the pendulum be by some means kept vibrating a long time and then released, the vibrations will die away more rapidly than before, as if the elasticity had become somewhat exhausted by prolonged exercise. This fatigue will persist for a long time but the wire will promptly recover after being heated to about 100° C.

181. Miscellaneous Properties of Solids. There are many mechanical properties of solids, frequently mentioned, which are not yet defined with sufficient clearness to make it possible to measure them but which call for some mention.

A **malleable** body is one which can be hammered into thin sheets. The most malleable substance is gold, which can be reduced to sheets of gold foil 1/250,000 inch in thickness.

A **ductile** substance is one which can be drawn out into fine wires. Silver and copper are very ductile; wires less than 1/1000 inch in diameter are readily made from these metals. By heating a substance until it is semiliquid and then drawing it out, fine threads of substances not ordinarily ductile can be made. Fine tubes and threads of glass are obtained in this way and fine threads of quartz, called quartz fibers, are thus made for use in suspensions of galvanometers and other instruments; they enabled Boys to greatly reduce Cavendish's apparatus (§ 146).

A **plastic** substance is one which can be moulded by pressure. Many substances not ordinarily regarded as plastic are so when subjected to great pressure slowly applied. A stick of sealing wax is ordinarily brittle but, suspended horizontally on end supports, it will slowly yield to its weight and bend. All metals under enormous shearing stresses become plastic. The impact of a cannon ball on armor-plate will sometimes produce a splash like a stone dropped in water.

A **friable** substance is one easily reduced to powder by a blow. Glass, diamond and crystals are friable.

Hardness is a term used in different senses. It sometimes means the opposite of plasticity, that is, resistance to change of shape, as when we speak of iron as hard and rubber as soft. Another use of it is to denote power of scratching, as in the mineralogists' scale of hardness, which con-

sists of a series of substances with diamond at one end, and talc at the other arranged so that each, beginning with diamond, will scratch the following but not the preceding. Any other substance that will scratch one in the list but not the next higher is said to have a hardness between the two.

PROPERTIES OF FLUIDS.

182. A **fluid** is distinguished from a solid by the absence of permanent resistance to forces tending to produce a change of shape; that is to say, *the shear modulus of a fluid is zero*. In this respect all fluids agree; they also agree in having weight and inertia. Because of agreement in these respects there are certain properties common to all fluids.

In certain other respects liquids and gases differ considerably. These differences are due to the fact that, while the particles of liquids are comparatively close together and attract each other with very considerable forces, the particles of gases are so far apart that the forces between them are negligible (except at impact). Properties in which liquids and gases differ will, therefore, be treated in separate chapters.

183. Direction of Force on the Surface of a Fluid. When a fluid is at rest the force acting on its surface must be perpendicular to the surface. This results from the fact that the shear modulus is zero; for, if the force were not at right angles to the surface, it might be resolved into a component perpendicular to the surface and a component along the surface. The latter would produce a sliding motion or a shear of the liquid near the surface so that the liquid could not be at rest.

At the surface of contact of a fluid and a solid, for example, at any part of the surface of a vessel in which the fluid is contained, the force exerted by the fluid on the solid is at right angles to the common surface. If it were not, the reaction of the solid on the fluid, being equal and opposite to the force of the fluid on the solid, would not be at right angles to the surface of the fluid.

At the free surface of a liquid, that is, where the liquid is in contact with a gas, the pressure between the two must be at right angles to the surface. The force of gravity must also be at

right angles to the free surface of a liquid at rest or sliding motion would result. Hence the free surface of a liquid at rest is horizontal unless it is acted on by some other force than gravity and gas pressure, such as surface tension (which we shall consider later) or magnetic force acting on a magnetic liquid.

184. Pressure in a Fluid. In a fluid there are forces, actions and reactions between contiguous parts of the fluid. These forces are due to several causes. The *weight* of the upper layers of the fluid has to be sustained by the lower layers and a pressure thus results. Force on the surface of the fluid, if it be completely enclosed, produces a pressure in the fluid; this is true not only of a fluid in a vessel which it completely fills but also of a liquid the free surface of which receives the pressure of the atmosphere or any gas above the liquid. (Another cause of pressure in a fluid is referred to in § 206.)

The total force exerted by a fluid on any surface is called the *thrust* on that surface. The thrust per unit of area at a point on the surface is called the *pressure intensity* or simply the *pressure* at that point. The pressure over a surface may be either uniform, that is, the same at every point, or variable. When uniform the pressure at any point equals the force on any unit of area; when variable it equals the average pressure, that is, the force on an area divided by the area, when the area is reduced without limit.

Whatever the causes of pressure in a fluid the *pressure at a point is the same in all directions,* that is to say, if we suppose an imaginary surface to separate the fluid at a point into two parts, the pressure of each part on the other is, as we have already seen, perpendicular to this surface, and it is also the same no matter how the imaginary surface is supposed to be inclined. This is nearly obvious from the mobility of the fluid, but the rigorous proof of the statement is not difficult.

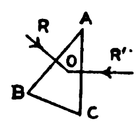


FIG. 83.

Let O be the point considered and let RO and $R'O$ be any two directions through O. Around O suppose a small prism described, and let two of its faces, of which AB and AC are the traces, be perpendicular to RO and $R'O$ respectively; while the third face, of which BC is the trace, is equally inclined to AB and AC, and let the ends of the prism be planes parallel to ABC. The fluid within the prism is at rest and therefore (neglecting its weight for a reason stated later) the

thrusts on all its faces form a system of forces in equilibrium. Hence the sum of the components of the forces in the direction BC equals zero. Two only of the thrusts have components in the direction BC, namely, those on AB and AC. Let these be R and R' respectively. They are equally inclined to BC, and if each makes with BC the acute angle θ,

$$R \cos \theta - R' \cos \theta = 0.$$

Now the areas of the faces AB and AC are equal; suppose each is a. Cancelling $\cos \theta$ and dividing by a we get

$$R/a = R'/a.$$

If we now suppose the prism to become indefinitely small, R/a becomes the pressure at O in the direction RO and R'/a becomes the pressure at O in the direction $R'O$. Since RO and $R'O$ stand for any directions through O, the pressure is the same in all directions.

As stated above the weight of the prism was neglected. As the prism is diminished without limit, the weight of the liquid in it, which is proportional to the cube of its dimensions, decreases more rapidly than the thrusts, which are proportional to the squares of the dimensions; each time the prism is reduced to one half in linear dimensions the area of each face is reduced to one fourth and the weight of the contained liquid is reduced to one eighth. Hence when the prism is taken small enough the weight becomes negligible compared with the thrusts.

185. Pressure at Different Points in a Fluid. (1) Let P and Q be two points in a fluid at rest, the positions of the points being such that the straight line PQ is horizontal and wholly in the fluid. Consider the forces acting on a cylinder of the fluid described about PQ as axis. The thrusts on the curved surface of the cylinder have no components in the direction of the axis. Hence for equilibrium the thrusts on the ends must be equal and opposite; and, since the ends are of the same area, the average pressures on the ends must be equal. If, now, the radius of the cylinder be supposed indefinitely decreased, the average pressures on the ends become the pressures at P and Q, which must, therefore, be equal. Hence the pressure in any direction at P equals the pressure in any direction at Q.

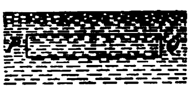


FIG. 84.

(2) Let P and Q be two points in a vertical line wholly in a fluid of density ρ. Consider the forces acting vertically on a

cylinder described with PQ as axis and of unit cross-section 1 sq. cm. If the depth of Q below P be h cms. the volume of the cylinder will be h c.c., its mass will be $h\rho$ gms. and its weight $h\rho g$ dynes. If p_1 be the pressure in dynes per sq. cm. at P and p_2 that at Q, the thrust downward at P will be p_1 and that upward at Q will be p_2. Hence

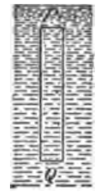


FIG. 85.

$$p_2 - p_1 = h\rho g$$

(3) Let P and Q be any two points in the fluid. No matter what the shape of the containing vessel, P and Q can be connected by a broken line made up of vertical and horizontal steps. Proceeding from P to Q there will be a difference of pressure $h'\rho g$ for each vertical step of length h'; while for each horizontal step there will be no change of pressure. Hence the total difference of pressure between P and Q will be $g\rho \times$ (the algebraic sum of the vertical steps) or, if the difference of level of P and Q be h, the difference of pressure will be $h\rho g$.

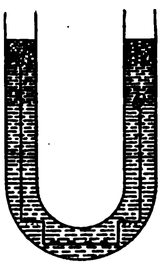


FIG. 86.

186. Pressure in a Gas. Since the density of a gas is comparatively small, the difference of pressure at two points is usually so slight as to be negligible; but this is not the case if h be very great. Thus in a vessel containing gas the pressure may be regarded as everywhere the same; but the pressure of the air varies greatly as we ascend to great heights in the atmosphere or descend to great depths in a mine.

187. Units Employed in Calculating Fluid Pressure. In establishing the formula for difference of pressure at different depths in a fluid, namely,

$$p_2 - p_1 = h\rho g$$

it has been supposed that absolute units are employed. If h be in cms., ρ in gms. per c.c. and g in cm. per sec.^{2} (about 980), p_1 and p_2 will be in dynes per sq. cm.

When the values of p_1 and p_2 would be inconveniently large in absolute units, other units may be employed. If g be omitted, p_1

and p_2 will be in gms. wt. per sq. cm. and

$$p_2 - p_1 = h\rho$$

This formula may also be used to calculate the pressure in metric tons (1,000,000 gms.) per sq. m. (10,000 sq. cm.) if h be in meters (100 cm.).

When British units are employed the weight of a cylinder of 1 sq. ft. cross-section and h feet in length and of density ρ (lbs. per cub. ft.) is $h\rho$ lbs. Hence, if p_1 and p_2 are in lbs. wt. per sq. ft.

$$p_2 - p_1 = h\rho$$

188. Surface of Contact of Two Fluids. The surface of contact of two fluids of different densities which are at rest and do not mix is horizontal. This may be deduced from the principle that for stable equilibrium the potential energy of a system must be a minimum (§ 107). If any part of the denser fluid were at a higher level than an equal part of the less dense, the potential energy could be decreased by interchanging the two. Hence, for the potential energy to be a minimum, every part of the denser fluid must be lower than any part of the less dense, that is, the surface of contact must be horizontal with the denser liquid below.

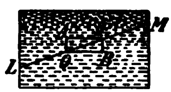


Fig. 87. The surface of contact of two fluids cannot be inclined as LM is.

Another proof is to suppose that the surface could be inclined, as LM. Let P and Q be two points in the surface. Complete a rectangle $AQBP$ with vertical and horizontal sides. The pressure at A would equal that at P and the pressure at Q would equal that at B. The increase of pressure from A to Q would equal the increase from P to B and this could not be the case unless the liquids were of equal density.

A particular case of the above is the surface of contact of a liquid with the atmosphere or any gas; it must be horizontal.

In both proofs it has been assumed that gravity is the only force acting on the particles of the fluids; if any other force exist the surface may not be horizontal. In any case it is at right angles to the resultant force.

189. Pascal's Principle. When a fluid is at rest the difference

of pressure between two points depends only on the difference of level and the density (§ 185). Hence if the pressure at any point be increased there will be an equal increase of pressure at every point (provided the density does not change appreciably) or *pressure is equally transmitted in all directions.* This is Pascal's principle of the transmissibility of pressure.

Pascal's principle is not rigorously true for a compressible fluid, for pressure will produce a change of density of a compressible fluid. But the compressibility of liquids is so small that the principle is practically true for all liquids. Gases are much more compressible, but their densities under ordinary circumstances are so small that the pressure in a moderate volume is everywhere practically the same and the principle is practically true for gases also.

190. The Hydraulic Press. In the hydraulic press Pascal's principle is applied to obtain a great force by the exertion of a relatively small one. It consists of a large cylinder and piston (or

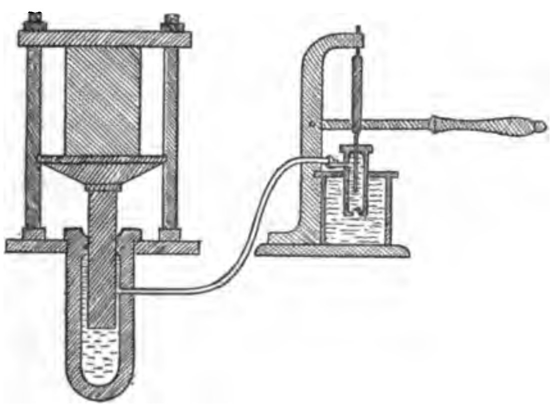


FIG. 88. Hydraulic press.

plunger) and a small cylinder and piston, the two cylinders being connected by a tube and filled with some liquid. Let the area of the large piston be S and that of the small one s. If the pressure in the liquid is p the thrusts on the pistons are pS and ps respectively, and these are in the ratio of $S : s$. Hence a small external force applied to the small piston will enable the large piston to exert a relatively great external force. The arrangement is indi-

cated in the figure; a valve in the connecting tube permits flow from the smaller cylinder toward the larger, but not in the opposite direction.

191. Archimedes' Principle. When a body is partly or wholly immersed in a fluid at rest, every part of the surface in contact with the fluid is pressed on by the latter, the pressure being greater on the parts more deeply immersed. The resultant of all these forces of pressure is an upward force called the *buoyancy* of the body immersed. The direct calculation of this resultant force is difficult except when the body has some simple form, such as a cylinder with its axis vertical; but a simple line of reasoning will show the magnitude and direction of the force.

The pressure on each part of the surface of the body is evidently independent of the material of which the body consists. So let us suppose the body, or as much of it as is immersed, to be replaced by fluid like the surrounding mass. This fluid will experience the pressures that acted on the immersed body and this fluid will be at rest; hence the resultant upward force on it will equal its weight and will act vertically upward through its center of gravity. It follows that *a body wholly or partly immersed in a fluid is buoyed up with a force which is equal to the weight of the volume of the fluid which the body displaces* and which acts vertically upward through the center of gravity of the fluid before its displacement. This point through which the force of buoyancy acts is called the *center of buoyancy*.

When a body floats partly immersed in a liquid, the weight of the body equals the weight of the liquid displaced. Let V be the volume of the body and D its density, v the volume of the part immersed and d the density of the liquid. Then $VDg = vdg$. Hence

$$\frac{v}{V} = \frac{D}{d}$$

or the fraction immersed equals the ratio of the density of the body to that of the liquid. If the density of an iceberg be taken as .918 and that of salt water as 1.026, the fraction of the iceberg below water is .918/1.026 or .895.

192. Fluids in Motion. While the calculation of the motion of a rigid solid body is comparatively simple owing to the fact that

we may treat a solid as a whole without regard to the actions between its parts, the discussion of the motion of a fluid is rendered difficult by the readiness with which any part of the fluid changes its shape, and we cannot, therefore, without the use of advanced mathematics, treat of any except a very few and simple cases of the motion of fluids.

When a fluid moves either in an open stream or in a closed pipe, the continual change of shape of each part is opposed by internal friction between these parts, and to maintain the motion some external force must be applied to the fluid. The most common causes of motions of fluids are gravity, as in the case of a river, pressure applied to some part of the boundary of the fluid, as in the case of water pumped through a system of piping, and the motion of solids in contact with the fluid, as in the case of a fan.

193. Motion of Fluids in Pipes. When the pressure on a fluid in a tube is greater at one end than at the other a flow ensues. When the pressure is first applied the motion begins with an acceleration, but after a time, if the pressures at the ends are kept constant and the supply of fluid is maintained, a steady state of motion ensues so that at each part of the tube the motion is constant. The simplest case is when the tube is of constant cross-section and the fluid is practically incompressible, that is, a liquid. In this case the volume of fluid passing all cross-sections of the pipe is the same throughout, and the rate of flow is, therefore, the same at each cross-section. The motion is from places of higher pressure to places of lower pressure. If, however, the fluid be compressible, while the motion at a point remains steady and the mass that passes every cross-section is necessarily the same as that which enters the pipe, the volume of flow is variable; for where the pressure is greater the fluid is compressed into a smaller volume, and where the pressure is less the fluid is not so much compressed. Thus the speed of the fluid particles is on the whole greater in the parts of the pipe where the pressure is less, that is, the further along the stream we go.

When a liquid flows through a tube of variable cross-section, the pressure at the ends being constant, the mass that passes each cross-section is the same but the rate of motion of the particles increases as the stream comes to a contraction of the tube, and

decreases again as the stream comes to an expansion of the tube. Now an increase of velocity or an acceleration necessarily means a smaller pressure ahead than behind, and a decrease of velocity necessarily means a larger pressure ahead than behind. Thus in a contraction (or "throat") the pressure is smaller than immediately before or behind, the amount of difference being dependent on the rate of flow through the tube and the cross-section at the throat and at either side. Similar considerations apply to the flow of gas through a pipe of variable cross-section, but this case is complicated by the changes of volume due to changes of pressure.

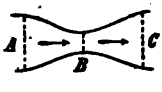


FIG. 89.

194. Examples of the Above. This principle is the basis of a common method of gauging the flow of water through pipes (the *Venturi meter*).

The adjacent figure represents (in section) a glass tube that passes tightly through a wide cork and a second cork through the center of which a pin is stuck. When air is blown through the tube the lower cork is not repelled but is attracted (the pin prevents side motion). The air increases its speed to pass through the small space between the corks, hence its pressure diminishes and atmospheric pressure pushes the corks together. Various other pieces of apparatus, such as the atomizer, the ball nozzle and the injector, act on the same principle.

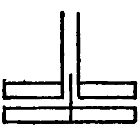


FIG. 90 a.

The curvature of the path of a rotating base ball or tennis ball is due to a difference of pressure on the opposite sides of the ball. Suppose the ball had no translatory motion but had a motion of rotation, while a current of air blew on it as indicated in the figure. The rotating ball would carry air around with it. At A the two air motions would conspire and at B they would be in opposition. Hence the velocity of the air-particles would be greater at A than at B and the pressure at B would, therefore, be greater than that at A, the result being a force on the ball in the direction BA. The same result follows when the ball is moving through air otherwise at rest, and the path curves toward the side of less pressure.

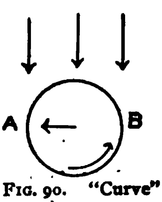


FIG. 90. "Curve" of a base ball.

195. Work Done by a Piston. When a piston of area a moves a distance d along a cylinder against a pressure p (per unit area), it exerts a force pa through a distance d, and therefore does

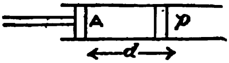


FIG. 91. Work by a piston.

work pad; and, since ad is the volume, v, through which the piston has moved, the work it does is pv.

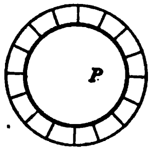


FIG. 92.

When a fluid contained in an extensible vessel (for example a gas in a rubber bag) increases in volume by an amount v work is done by the fluid. If the pressure p is kept constant (or if the expansion is very slight) the work done is pv. For each small part Δs of the surface of the vessel acts like a small piston of area Δs which moves through a small volume Δv. The whole work is the sum of $p\Delta v$ for all the small pistons and this sum is pv.

Conversely, to compress a fluid work must be done on it and the amount of work is pv, where p is the pressure (supposed constant) and v the decrease of volume.

196. Viscosity. A fluid offers no permanent resistance to forces tending to change its shape; it yields steadily to the smallest deforming force. But the rate of yielding is different for different fluids, that is, different fluids offer different *transient* resistances to deformation. This transient resistance is called internal friction or *viscosity*. Thus a very viscous liquid such as glycerine or tar flows much more slowly through a tube or down an incline than water does, and such a flow consists in a continuous change of shape of each part of the liquid. The internal forces are what we have called stresses, and, since the strain is a change of shape only, the stress must be a shearing stress.

Consider, as an example, the motion of a stream flowing down a very gentle incline under the force of gravity. The motion is greater near the surface than at the bottom. A small cube $ABCD$ with sides vertical and horizontal will, by the motion, be changed into the form $abcd$. The liquid above AB exerts a force in the direction AB, on the upper face of the cube, and the liquid below CD exerts a resisting force on the face CD in the direction CD. These two forces constitute a shearing stress. A similar description applies to a small cube of a liquid flowing in any manner whatever. Very extensive experiments have shown that *the ratio of the shearing stress to the rate of shear is a constant for any one fluid,* the value of the constant being different for different fluids. This is the fundamental and very simple law of fluid friction.

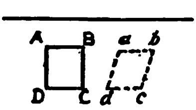


FIG. 93. Shearing of a fluid.

The constant ratio of the shearing stress in a fluid to its rate of shear is called the coefficient of viscosity of that fluid.

A concrete case will make this definition clearer and will lead to another very common way of stating the definition. Suppose that the space between two large parallel plates is filled with the fluid under consideration and let one plate be moving parallel to the other with a velocity v. Experiment (as stated later) shows that the fluid in contact with the plates does not slip along the faces of the plates but adheres to them. The moving plate will in a very short time t move a distance Vt, and, if the distance between the plates be d, the shear produced in the time t will be Vt/d. Hence the *rate of shear* is V/d. If the area of each plate be A and the force applied to move the upper plate be F, *the shearing stress* will be F/A. Hence, denoting the coefficient of viscosity by μ, we have

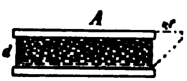


FIG. 94.

$$\mu = \frac{F/A}{V/d}$$

or

$$F = \mu \cdot \frac{AV}{d}$$

If now we suppose A, V and d to be each unity, F will be equal to μ. Hence we have the following definition of μ: *The coefficient of viscosity of a fluid is the tangential force on unit of area of either of two horizontal planes at the unit distance apart, one of which is fixed while the other moves with the unit velocity, the space between them being filled with the viscous material.*

197. Measurement of Coefficients of Viscosity. The most common method of finding μ is by measuring the flow of the fluid through a tube of very small bore (or so-called capillary tube). The motion of the fluid in such a case (provided the velocity does not exceed a certain magnitude) is analogous to the slipping of the tubes of a small pocket telescope through one another. If we imagine the fluid divided into a very large number of thin cylindrical shells, the motion consists of the slipping of shell through shell; hence the resistance encountered is internal friction or viscosity. Let p be the difference of pressure per unit area at the two ends of the tube (supposed horizontal), l the length of

the tube, and r its radius. It has been shown theoretically and experimentally that, when the fluid is a liquid, the volume that flows out of the tube in unit time is

$$V = \frac{p\pi r^4}{8l\mu}$$

This formula also applies to a gas, if p be very small, but if p be large the formula must be modified to allow for the compressibility of the gas. In the theoretical proof of the above formula it is assumed that no slipping of the fluid along the surface of the tube takes place, and the agreement of theory and experiment confirms this assumption.

The following are some values of μ in C.G.S. units at 20° C.

Alcohol	0.0011	Water	0.10
Ether	0.0026	Glycerin	8.0

198. The Explanation of Viscosity. Viscous resistance to fluid motion resembles friction between solids in certain respects, and in other respects the two are very different. Both are forces that appear only as resistances to relative motion; they are, therefore, non-conservative forces and energy spent in doing work against them is changed into heat. But, while the friction between solids is, through a considerable range of velocity, independent of the velocity, the resistance due to viscosity is exactly proportional to velocity through the widest range in which experimental tests have been made. This points to a fundamental difference in the nature of the resistance in the two cases.

There are many strong reasons for believing that the particles of fluid are in rapid motion and are not, like the particles of solids, confined to certain positions. If now we imagine two layers of a fluid in relative motion, so that one is passing another, like one railway train passing a second, it is evident that particles from each layer must be continually crossing the boundary into the other layer. The particles of the more rapidly moving layer that cross the boundary carry their larger momentum with them and thus produce a gradual increase of the velocity of the second layer. At the same time particles of the latter layer penetrate into the former and by taking up momentum diminish the velocity of that layer. The result, on the whole, is a tendency of the two layers to come to the same velocity, and this is exactly what we mean by a resistance to relative motion. In the case of gases this explanation may be regarded as fully established; for the formulas to which it leads by mathematical methods are verified by experiment. It has not yet been found possible to work out the mathematical results in the case of liquids, but there is no reason to doubt that the explanation is equally applicable to the latter.

Liquids.

199. Compressibility of Liquids. While the shear-modulus of any liquid is zero the bulk-modulus is usually large, that is, the pressure on a liquid must be greatly increased to produce much diminution of volume. The coefficient of compressibility of a liquid (§ 169) is therefore small. Measurements of the compressibilities of liquid are made by subjecting the liquids to great pressures in a vessel called a piezometer and noting the resulting diminution of volume.

The following table gives the compressibilities of some liquids. Each number is the proportion by which the volume of the liquid is decreased when the pressure on it is increased by one atmosphere.

Alcohol	0.0000828	Mercury	0.0000038
Ether	0.0001156	Water	0.0000489

200. Hydrometers. A Hydrometer is an instrument for finding the density of liquids; some hydrometers may also be used to find the density of solids. The action of most hydrometers depends on Archimedes' principle. Some hydrometers sink to different depths in different liquids and thus indicate the densities of the liquids; these are called hydrometers of *variable immersion.* Other hydrometers are used with different weights added to the weight of the instrument so that they are always immersed to the same depth; these are called hydrometers of *constant immersion.*

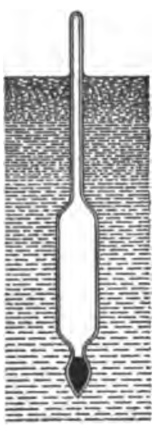


Fig. 95. Common hydrometer.

The common hydrometer is one of variable immersion. It is a glass tube with an enlargement in the middle and weighted at the lower end with mercury so that it will float in stable equilibrium. Inside the tube is a scale which indicates the density of the liquid by the depth to which the tube is immersed.

The best known hydrometer of constant immersion is Nicholson's Hydrometer. It consists of a hollow cylindrical body (of metal or glass), to one end of which a somewhat heavy basket *B* is attached, while at the other end there is a stem *S* which carries a scale pan *C* for weights.

On the stem there is a mark indicating the depth to which the hydrometer is to be immersed. Let W be the weight of the hydrometer and let w be the weight that must be placed on the pan to make the instrument sink to the mark in water of density d, and w' the weight on C required when the hydrometer is in a liquid of density D. The volume of liquid displaced in both cases is the same. Hence, by Archimedes' principle, the weights of equal volumes of the second liquid and of water are $W + w'$ and $W + w$. Hence

$$\frac{D}{d} = \frac{W + w'}{W + w}.$$

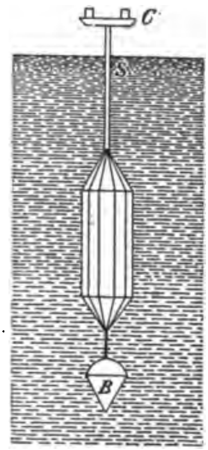


FIG. 96. Nicholson's hydrometer.

This hydrometer may also be used to find the density of a small solid. When so used the instrument is in reality a balance for weighing the solid in air and then in some liquid of known density. The body is first placed on C. The weight required on C to sink the hydrometer to the mark on the stem will be less than w' by the weight of the body. This gives the weight of the body in air. The body is then placed in B and its apparent weight when immersed is found in the same way. The ratio of the weight of the body to its apparent loss of weight when immersed, which equals the weight of an equal volume of liquid, gives the specific gravity of the body relatively to the liquid.

201. Stability of Flotation. A body floating at rest on the surface of a liquid is in equilibrium under the action of its weight acting vertically downward through the center of gravity, G, and the resultant upward pressure of the liquid acting through the center of buoyancy, B. Hence the two forces are equal and act in opposite directions in the vertical line BG. Suppose the body to rotate slightly about an axis perpendicular to the plane represented in the figure. The form of the volume of water displaced is now different (unless the body be spherical or cylindrical) and the center of buoyancy is at some point B' not in the vertical line through G. Hence the forces now acting on the body constitute a couple and if the couple tends to right the body the equilibrium is stable, if not it is unstable.

The simplest case to consider is when the body is symmetrical, or very nearly so, on opposite sides of the plane through B and G perpendicular

to the axis of rotation; for in this case B' is in this plane. Let a vertical through B' cut BG in M. For small rotations the position of M on BG is very nearly independent of the magnitude of the rotation. M is called the metacenter of the body. The position of M can usually be calculated by mathematical methods. If M is above G it is evident that the couple tends to right the body and the equilibrium is stable; if M is below G the

Fig. 97. Stable equilibrium of a vessel.

couple tends to displace the body further and the equilibrium is unstable. Hence the danger of taking the whole cargo out of a vessel without putting in ballast and the risk of upsetting when several people stand up at once in a small boat. A ship has one metacenter for rolling and another for pitching. In general the vessel is not quite of the symmetrical form assumed above and the problem of stability is more complicated.

202. Energy of a Moving Stream. When liquid flows steadily through a pipe of varying cross-section, the total energy in the space between any two sections A and B remains constant. When a volume V flows in through A, an equal volume flows out through B. Let the pressure at A and B respectively be p_1 and p_2, and the velocities v_1 and v_2 respectively. Let ρ be the density of the liquid. When the volume V flows in through A it carries kinetic energy $\frac{1}{2}V\rho v_1^2$ into the space between A and B and in the same time the volume V flows out through B carrying energy $\frac{1}{2}V\rho v_2^2$.

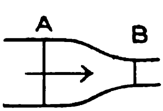


Fig. 98.

Now the liquid to the left of A acts like a piston in forcing liquid into the space between A and B, and it thus does work p_1V which goes to increase the energy between A and B; and in the same time the liquid between A and B does work p_2V in forcing liquid out through B. From this cause there is an increase of energy $p_1V - p_2V$ between A and B. If between A and B there is a fall of level from h_1 to h_2, the liquid which flows in at A will have a greater amount of potential energy than that which flows out at B, and there will, therefore, be an increase of potential energy of $V\rho g(h_1 - h_2)$ between A and B. But the total energy between A and B remains constant. Hence

$$\tfrac{1}{2}V\rho(v_1^2 - v_2^2) + (p_1V - p_2V) + V\rho g(h_1 - h_2) = 0$$

or

$$p_1 + g\rho h_1 + \tfrac{1}{2}\rho v_1^2 = p_2 + g\rho h_2 + \tfrac{1}{2}\rho v_2^2 = \text{a constant}$$

This is Bernoulli's theorem. It is of fundamental importance in hydraulics.

203. Outflow from an Orifice. Torricelli's Theorem. When an orifice is opened in a side of a vessel containing liquid at greater than atmospheric pressure, the liquid is forced outward. The simplest way of finding the velocity of the escaping liquid is by an application of the principle of conservation of energy.

A small mass m of liquid escaping with velocity v has $\frac{1}{2}mv^2$ units of kinetic energy. If no liquid has been added to the vessel during the escape of the mass m, the potential energy of the liquid in the vessel must have diminished by an amount equal to $\frac{1}{2}mv^2$. The mass m was really removed from the part of the liquid near the orifice, but the change of the state of the liquid in the vessel is the same as if the mass m had been removed from the surface; and the change of total potential energy of the liquid in the vessel and of the escaping liquid is the same as if a mass m had been lowered from the surface to the depth of the orifice. Hence, denoting the depth of the orifice below the surface by h, the loss of potential energy is mgh and, therefore,

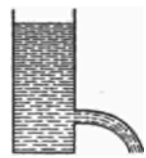


FIG. 99.

$$\tfrac{1}{2}mv^2 = mgh$$

and

$$v = \sqrt{2gh}$$

Thus *the velocity of escape is the same as if the escaping liquid had fallen freely through the distance of the orifice below the surface.* This is known as Torricelli's Theorem. It was first stated by a pupil of Galileo named Torricelli, who also discovered the principle of the barometer.

Torricelli's Theorem may also be deduced from Bernoulli's theorem but we shall leave the deduction as an exercise for the reader.

The above theorem relates only to the velocity of the particles as they leave the orifice. It does not enable us at once to calculate the volume that escapes in a given time; for the cross-section of the jet contracts for a short distance after leaving the vessel, and at a certain point reaches a minimum called the *vena contracta* (or contracted vein) beyond which it expands. If the area a of the cross-section of the *vena contracta* is found, the volume per second that escapes is av. The ratio of a to the area of the

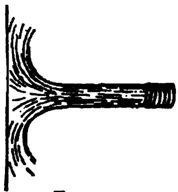


Fig. 100.

orifice depends on the velocity of escape and can be changed by inserting a tube (or *ajutage*) through the orifice.

204. Pressure Exerted by a Stream. When a stream of liquid meets an obstacle and is arrested it gives up its momentum to the obstacle, that is, it exerts a force on the obstacle. The pressure thus produced can be calculated from the velocity of the water and the amount of water that impinges per second on this obstacle. On this is founded a method of measuring the velocity of a stream. A tube bent at right angles is placed in the stream so that one arm points horizontally up stream and the other vertically upward. It the water were at rest the liquid would rise in the vertical axis to the height of the surface of the water, but the pressure of the stream raises it higher and from this additional height the velocity of the stream can be deduced.

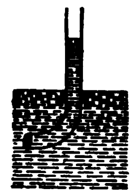


Fig. 101. Pitot tube for finding velocity of stream.

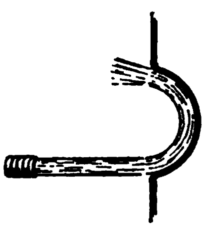


Fig. 102.

When a jet impinges on an obstacle and flows off laterally the pressure exerted is that due to the loss of the momentum of the liquid. If this obstacle is curved so that the motion of the liquid is reversed the water is given an equal momentum in the opposite direction and the force exerted on the obstacle is doubled. This principle is taken advantage of in the construction of water-wheels.

When a stream strikes an obstacle obliquely it is partly arrested and then flows down along the surface of the obstacle. Thus the side of the obstacle farther up stream receives more momentum than the lower side and so tends to turn more nearly perpendicular to the stream. A floating log free only to swing about its middle point sets itself across the stream. A leaf falling from a tree tends to take a horizontal position. The effect is readily illustrated by sweeping through the

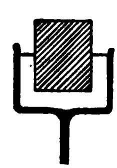


Fig. 103. A disk swept through the air turns perpendicular to the direction of motion.

air a square disk of cardboard which is connected by short threads to a wire frame.

205. The Hydraulic Ram. Water flowing under the action of gravity tends to the condition in which it would be in equilibrium, and in which, therefore, all parts of the free surface would be at the same level. This is the meaning of the statement that "water seeks its own level." Usually it is only by the means of work done by some force other than gravity that water can be raised to a higher level. In the hydraulic ram a small fraction of the water in a stream is raised to a high level by a self-acting mechanism which does not need any external power.

When a stream of water in a pipe is suddenly stopped, for example when a water-tap is turned off, the momentum of the water, which may be very large, is stopped in a very short time and therefore the force exerted by the water on the pipes may be very much larger than that which the water exerts after it has come to rest. In the hydraulic ram this momentary intense pressure is used to drive water into an air-chamber such as is used in a force-pump.

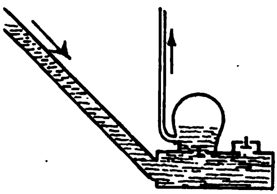


Fig. 104. Principle of hydraulic ram.

Momentary interruptions of the current are caused by the opening and closing of a valve which works automatically in a vertical direction. The weight of the valve is such that, when it is closed and the water is at rest, the pressure of the water on the lower surface of the valve is not sufficient to keep it closed; hence it opens and allows the stream to start. The stream when in motion carries the valve with it, again closing it and arresting the motion.

Some of the potential energy of the head of water is transformed into kinetic energy of the flowing stream, and this is partly changed into potential energy of the compressed air, which again is changed into potential energy of the water at the top of the delivery tube. Only a small part of the water is finally raised to a higher level than its original one, and its gain of potential energy is compensated by the loss of potential energy of the remainder.

Molecular Properties of Liquids.

206. Molecular Forces. Between the particles of a solid or of a liquid there are attractions that keep the body together unless these forces are overcome by external forces. To show directly the existence of these forces between the particles of a liquid is very difficult, since a liquid so readily changes its shape. It has, however, been found possible to fill a glass tube with water at a high temperature and then seal the tube; the water, on cooling, continued to fill the tube, without contracting, until it exerted a tensile force of over seventy pounds per square inch upon the walls of the tube. The water would, in such an experiment, stand a much higher stress if it were possible to free it perfectly from absorbed gases. It is this attraction between the particles of a liquid that has to be overcome when a liquid is evaporated; and, from the heat required for evaporation, it can be calculated that the attractions between the particles are very powerful and produce a very great internal pressure across any imaginary plane in the liquid.

From the above it might be thought that a body immersed in a liquid would feel the effect of this great internal pressure. That such is not the case is due to the fact that the molecular forces of attraction are sensible only when the distances between the particles are exceedingly small. Thus two particles of water practically cease to attract when the distance between them exceeds a value that is, roughly, about .00005 mm. Now the thinnest solid that it is possible to insert in a liquid separates the particles so far that the attractions between them are negligible, and thus the pressure on an immersed solid is merely that due to the causes, gravitation and pressure on the boundary, considered earlier.

The distance to which the force of attraction is sensible is called the *range of molecular forces.* Any particle of a liquid is attracted by all particles that lie within this range and these are contained within a sphere. This sphere, whose radius is the range of molecular forces, may be called the *sphere of influence* of a particle.

207. Surface Tension. The molecular forces of which we have been speaking produce certain remarkable effects at the surface of a liquid. *The surface of a liquid tends to contract to the smallest area admissible.* Thus a drop of water falling through

the air becomes spherical, since the sphere is the figure of least surface for a given volume. The same is true of a drop of liquid lead falling in a shot tower; the drop solidifies during the fall and is found to be spherical when the fall is sufficient to allow it to become perfectly solid while in the air. A mixture of alcohol and water can be prepared of the same density as an oil, and a large drop of the mixture floating totally immersed in the oil is spherical. When the end of a stick of sealing-wax or of a glass rod is melted in a flame, it tends to the spherical form. A beautiful illustration of the tendency of a liquid surface to contract consists in forming a film from a soap-bubble solution on a ring of wire, to which a loop of silk has been loosely attached so that the loop floats in the film; when the film is broken inside the loop the latter becomes circular. In shrinking to the form of least area the film pulls the loop into the form of greatest area for a given periphery, and this is a circle.

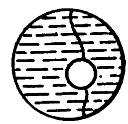


FIG. 105. Loop of silk on surface of film.

208. Explanation of Surface Tension. Consider the condition of a particle at *A* in the body of a liquid, and that of a particle at *B*, at less than the range of molecular forces from the surface. The particle at *A* is equally attracted on all sides by the particles around it, but the particle at *B* is more attracted inward than outward, since a sphere with center at *B* and the range of molecular forces as radius lies partly outside of the liquid. To take a particle from *A* to *B*, work must be done against this inward attraction.

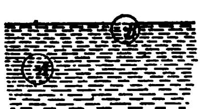


FIG. 106.

Now, when the surface of a liquid is increased, for example when a soap film is stretched, more particles are drawn into the surface; hence some work is done by the stretching force and therefore an opposing force is overcome. But the stretching force required is parallel to the surface; hence the liquid exerts an opposing or contractile force parallel to the plane of the surface, and this force is what we call the surface tension. Thus we explain the existence of a tension in the surface of a liquid by showing that it is in accordance with the principle of work. At present our knowledge of the state of the particles near the surface is too

imperfect to enable us to describe their condition more precisely and to show how the state of tension along the surface is produced.

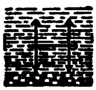


FIG. 107. Surface tension is the force across unit length.

If a line be imagined drawn along the surface of a liquid, the part of the surface on one side of the line pulls on the part on the other side, and if the length of the line be supposed one cm. the pull in dynes is taken as the magnitude of the surface tension of the liquid; this we shall denote by T.

209. Methods of Measuring Surface Tension. Surface tension manifests itself in many ways and, as almost any of its effects may be made the basis of a method of measuring it, the methods that have been employed are numerous. When the liquid can be formed into a thin sheet, as in the case of a soap solution, a direct method of measuring it may be used; a film may be formed on a wire frame of which one side is movable; if the force required to hold this side at rest against the surface tension is F, and the length of the movable side is l, the tension in each surface of the film is $F/2l$.

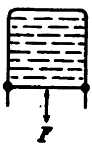


FIG. 108. Stretching a film.

To draw a horizontal wire up through the surface of a liquid the tension of the surface must be overcome, and from the force required the surface tension may be calculated.

The movement of minute waves or ripples on the surface of a liquid is due chiefly to the surface tension of the liquid, and from the wave-lengths of the ripples and their velocities we can find the magnitude of the surface tension.

The rise of liquid in a capillary tube depends, as we shall see later, on the surface tension of the liquid, and this affords another method of measurement.

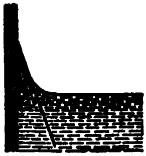


FIG. 109. Contact of water and glass.

210. Contact of Liquid and Solid. The general free surface of a liquid is horizontal; but, where the liquid is in contact with a solid, the surface is usually curved, the direction and amount of the curvature being different for different liquids and different solids. Water in contact with a vertical surface of glass is curved upward, and mercury in the same circum-

stances is curved downward. These, for a reason stated later, are called *capillary* phenomena.

The contact angle of the wedge-shaped part of the liquid between the free surface of the liquid and the surface of the solid is called the *angle of capillarity*. The size of the angle in any case depends on the purity of the liquid and the cleanness of the solid surface. Thus for very pure water in contact with clean glass the angle is 0°; but with slight contamination, even such as is caused by exposure to air, the angle may become as large as 25° or more. For perfectly pure mercury and glass the angle is about 148°, but slight contamination reduces it to 140° or less; for turpentine it is 17°, for petroleum 26° and so on.

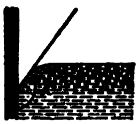


FIG. 110. Contact of mercury and glass.

211. Level of Liquids in Capillary Tubes. When a glass tube of very fine bore (or so-called capillary tube), open at both ends, is placed vertically with its lower end in a vessel of liquid, the surface of the liquid in the tube is usually higher or lower than the general level of the surface in the vessel. When the liquid is water or alcohol the surface is elevated in the tube; when the liquid is mercury the surface is depressed. For a given liquid the amount of elevation or depression is greater the smaller the bore of the tube, being, in fact, inversely as the diameter of the bore. For tubes of other materials than glass similar effects, depending in amount on the material of the tube, are observed.

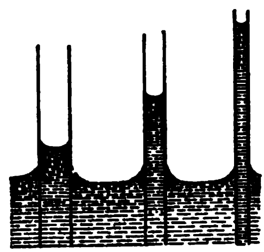


FIG. 111. Water in capillary tubes.

There are similar elevations and depressions between two glass plates standing close together in a liquid. These elevations and depressions and the curvature of a liquid surface in contact with a solid are usually grouped under the general title of *Capillarity*.

Assuming the existence of the invariable angle of capillarity at which a liquid meets a solid, we can give a simple explanation of capillary elevations and depressions.

Consider the case when the liquid is elevated. The liquid in the tube meets the tube in a circle of radius r equal to the radius of the bore, and at every point of the circle the angle of contact

is the angle of capillarity α. Thus the surface tension of the liquid pulls on the tube in the direction PQ inclined at α to the length of the tube; and the tube therefore reacts with an equal pull in the direction QP. The amount of the pull per unit length of the circumference of the circle of contact is T, and the component of this, parallel to the length of the tube, is $T \cos \alpha$. For the whole circumference of the circle of contact the sum of these components is $2\pi rT \cos \alpha$. This is an upward force on the liquid in the tube, and it draws the liquid upward until the weight of the liquid elevated above the ordinary surface equals the supporting force. If the mean elevation is h, the volume of the supported column is $\pi r^2 h$ and its weight $\pi r^2 h\rho g$ in dynes. Hence

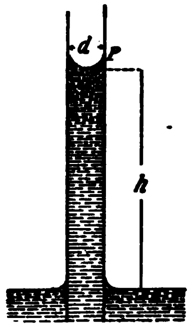


FIG. 112.

$$\pi r^2 h\rho g = 2\pi r T \cos \alpha$$

$$\therefore h = \frac{2T \cos \alpha}{g\rho r}$$

Thus the elevation is directly as the surface tension and inversely as the radius of the tube. By measuring the elevation and the radius and finding α by some other method, the value of T for any liquid may be obtained.

212. Elevation between Plates. The above method of proof may also be extended to the case of a liquid between parallel plates (Fig. 112). In this case the surface of the liquid meets the surfaces of the plates in straight lines. Let the distance between the plates be d. Consider the equilibrium of the liquid contained between the plates and two vertical planes perpendicular to the plates and at unit distance apart. The pull of the surface tension at the top is $2T \cos \alpha$ and the weight of the liquid supported is $dh\rho g$. Hence

$$h = \frac{2T \cos \alpha}{g\rho d}$$

Thus the elevation is the same for two parallel plates as for a tube if the distance between the plates equals the radius of the tube.

213. Pressure Caused by a Curved Surface under Tension. Since the liquid in a capillary tube is elevated above or depressed

below the ordinary level, the pressure beneath the curved surface must be less or greater than the pressure at the general surface. When the effect is a depression (mercury is glass), the depressed surface is curved downward and the tension in the surface produces a pressure, just as the tension in a rubber sheet stretched over a ball produces pressure on the ball. When the effect is an elevation, the stretch on the upward curved surface tends to draw the liquid in the surface layer away from the liquid below and so produces a state of tension or diminution of pressure beneath the surface. From the amount of the elevation or depression we can calculate the change of pressure thus caused. In the case of an elevation to a height h the pressure must be less than the pressure at the ordinary level, which is atmospheric pressure, by $g\rho h$ or (§ 211) $(2T \cos \alpha)/r$. Here r is the radius of the tube. If we denote the radius of the spherical surface by R, $R \cos \alpha = r$. Hence the pressure beneath the concave surface is less than that of the atmosphere above by $2T/R$. The same applies to the pressure produced on the concave side of a depressed surface. This difference of pressure on the two sides is due entirely to the tension and the curvature of the surface.

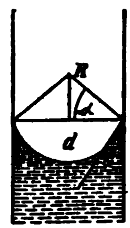


FIG. 113.

In the case of a spherical soap-bubble there are two surface tensions to be considered, one on the inner side of the film and the other on the outer side. Hence the total pressure inside the bubble, due to the tension and curvature of the film, is $4T/R$.

A cylindrical surface in a state of tension also produces pressure on the concave side. This is deduced, as above, from the elevation or depression of a liquid of surface tension T between two parallel plates at a distance d apart. If R is the radius of the cylindrical surface of the liquid $R \cos \alpha = \frac{1}{2}d$. Hence (§ 211) $p = T/R$, and this is therefore the pressure on the concave side due to the tension T in a cylindrical surface of radius R. In the case of a cylindrical soap-bubble of radius R the tension in each surface produces pressure T/R. Hence the pressure inside is greater than that outside, by $2T/R$.

214. Other Effects of Surface Tension. When the angle of capillarity of a liquid in contact with a solid is small, the liquid, in its attempt to establish this small angle, spreads out on the surface of the solid; that is,

the liquid is one that wets the solid. Thus a drop of water let fall on clean glass spreads out, the angle of capillarity being small. A drop of mercury on a glass plate has no tendency to spread but gathers into a ball.

FIG. 114. Water between glass plates.

A film of water between two glass plates makes it difficult to draw the plates apart by a force normal to their surfaces. The liquid tends to spread over both plates and become concave outwards, so that the pressure within it is less than the atmospheric pressure which acts on the outside of the plates, and this produces an apparent attraction between the plates.

When an attempt is made to blow out a glass tube containing numerous detached drops a surprising resistance is experienced. Each drop becomes concave on the side of high pressure and the total resistance is the sum of the pressures exerted by these concave surfaces.

Small bodies, such as straws and sticks, floating on the surface of a liquid usually attract and gather into groups. Let us represent two such bodies by small vertical plates. If the liquid wets both it rises between

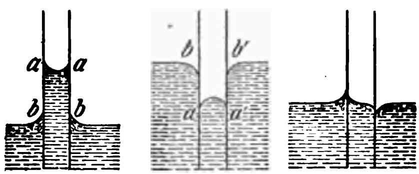


FIG. 115. Capillary attractions and repulsions.

them, and the pressure in the elevated portion is less than the atmospheric pressure on the outer sides of the plates. Hence the plates are pushed together. If the liquid does not wet either plate it is depressed between them; the pressure above the depressed part is atmospheric, while the pressure in the liquid on the outer sides of the plates is greater than atmospheric and the plates are pushed together. If the liquid wets one plate but not the other there is a part of each plate on which the pressure on the inside is greater than that on the outside; hence an apparent repulsion results. (Balls of paraffine wax some of which are lamp-blacked, floating on water, will illustrate all three cases.)

Any dissolved substance or impurity weakens the surface tension of water. This explains the irregular motions of small particles of camphor dropped on clean water. At some points the camphor dissolves more rapidly than at other points, and near the former the surface tension of the water is weakened so that the pull on the opposite side, where the tension is greater, prevails and causes irregular motion.

215. Diffusion of Liquids. The gradual mixture of two liquids which come into contact is called *diffusion*. It takes place on a

large scale where fresh water from a river flows out into the ocean. It may be illustrated on a small scale by pouring a solution of a colored salt into a tall vessel and then cautiously covering the colored solution with a layer of water. The particles of each liquid are in motion and begin to make their way across the interface, and, after a long time, the whole vessel is filled with a mixture of the same constitution throughout. Stirring has the effect of increasing the area of contact of the liquids and so promotes diffusion. The liquids must be such as will "mix"; oil and water, not being capable of mixing, will not diffuse across a surface of contact.

Let us denote the two diffusing liquids by A and B, and let us suppose that initially A occupies the lower half of a tall jar and B the upper half. The concentration of either of the liquids at any point is its mass per unit volume at that point (*i. e.*, its density at the point if the other liquid be imagined absent without the first being disturbed). The liquid A diffuses vertically upwards, that is, from places of high concentration to places of low concentration. The *gradient of concentration* in any direction is the rate at which the concentration falls off in that direction; if the rate of fall per unit of distance is unity, the gradient of concentration is unity. The general law of diffusion is that *the rate of diffusion for each liquid is proportional to the gradient of concentration* of that liquid. The *coefficient of diffusion* or the *diffusivity* of the liquid is the mass in grams that crosses unit area in a day when the gradient of concentration is unity. This constant can be found from observations of the density at various points along the direction of diffusion, made by means of beads of different densities floating in the liquid, and in various other ways. The following table contains the coefficients of diffusion of various substances into water at the temperature (Centigrade) stated.

Hydrochloric acid	1.74	at	5°
Common salt	0.76	"	5°
Common salt	0.91	"	10°
Sugar	0.31	"	9°
Albumen	0.06	"	13°
Caramel	0.05	"	10°

From the above it will be seen that liquids vary widely in diffusivities. Substances of high diffusivity are called *crystalloids* and those of low diffusivity are called *colloids*. The former group includes mineral acids, salts and substances generally that form crystals (whence the name), while the latter includes gums, albumens, starch, and glue (the name being derived from the Greek for glue). Crystalloids dissolved in water produce many marked changes in its properties; colloids in water form jellies, which seem to consist of a semi-solid framework holding the liquid in its meshes. Colloids have large and complex molecules and it is perhaps to this fact and to the consequent slower motions of the molecules that their small diffusivities are due. They are comparatively tasteless, as they do not diffuse and reach the nerve terminals. Their low rates of diffusion also render them indigestible. Through a layer of a colloidal jelly crystalloids will diffuse almost as rapidly as through water, but colloids not at all.

216. Diffusion through Membrane. Osmosis. Through certain membranes which have no visible pores, many liquids will diffuse readily. Thus through a partition of rubber between water and alcohol the alcohol will pass rapidly, while the passage of the water is barred. If animal membranes are wet by water it readily passes through. A method of separating crystalloids and colloids, called dialysis, depends on the different rates at which these substances pass through such a membrane as parchment paper. The diffusion of substances through such septa is called *osmosis*.

Some membranes allow one constituent of a mixed liquid or solution to pass, while barring the other constituent; such membranes are called *semi-permeable*. One such is ferrocyanide of copper, formed in the pores of a porous partition by the reaction between ferrocyanide of potassium on one side and copper sulphate on the other. When such a membrane separates water and the aqueous solution of any one of various salts, the salt does not pass but the water passes in both directions, though more rapidly towards the solution than in the opposite direction. If the solution be in a tube the lower end of which, closed by a plug of the membrane, is dipped in water, the level in the tube will rise until

(provided the membrane does not break) the column is of such a height that its pressure prevents further flow. This pressure is called the *osmotic pressure* of the solution. Its magnitude for very weak solutions is proportional to the concentration, that is to the number of molecules of the dissolved salt per unit volume. For a large number of salts the pressure is the same for solutions that contain the same number of molecules of the salt in unit volume. For various other salts the osmotic pressure for a given number of molecules per unit volume is two (or some whole number of) times greater than for the first group; this is possibly due to the molecules being resolved into atoms in the solution, the atoms acting independently. But the full explanation of osmosis and osmotic pressure is a matter of much dispute. One remarkable fact may be noted, namely, that the osmotic pressure for a given number of molecules (or of dissociated atoms) in an aqueous solution is equal to the pressure that these molecules (or atoms) would produce if freely flying as gaseous particles in the space occupied by the solution. It is also noteworthy that the osmotic pressure increases in the same way and at the same rate with rise of temperature as the pressure of a gas does.

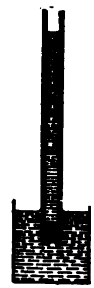


FIG. 116. Osmotic pressure.

Osmosis plays an important part in many natural processes.

PROPERTIES OF GASES.

217. A gas has already been defined as a fluid which has no definite volume of its own independent of the containing vessel, but expands so as to occupy any vessel in which it is contained. Gases have the same properties as liquids in all respects which depend on the fact that the shear modulus of a fluid is zero. The pressure at a point in a gas is the same in all directions (§ 184). The pressure of a gas on a surface is normal to the surface (§183). Pressure applied to any part of the boundary is equally transmitted in all directions (Pascal's Principle § 189). A body im-

mersed in a gas is buoyed up with a force equal to the weight of the gas displaced (Archimedes' Principle § 191). The pressure in a gas increases with its depth at a rate expressed by $g\rho h$, as in the case of liquids (§ 185). Gases also show the property of internal friction or viscosity, and the definition of the coefficient of viscosity of a gas is the same as that of a liquid. Some of these properties are of special importance in the case of a gas and call for separate treatment.

218. Pressure of the Atmosphere. A very important example of the pressure of a gas is the pressure exerted by the earth's atmosphere. The atmosphere, consisting chiefly of oxygen and nitrogen, is held to the earth by the gravitational attraction between it and the earth. The total pressure on the surface of the earth is the total attraction between the earth and the atmosphere, that is, the weight of the atmosphere. The pressure on any horizontal area of the earth's surface is the weight of all the air vertically above that area. At the top of a mountain the pressure is less than at sea level, since less of the atmosphere is above.

Fig. 117. Torricelli's experiment.

Galileo discovered that air had weight by weighing a glass globe containing air and then re-weighing it when he had forced more air into it. His friend and pupil Torricelli found (in 1643) that, when a tube 33 inches long filled with mercury and closed at one end was inverted in a dish of mercury, the mercury stood at a height of about 30 inches in the tube,, thus leaving a vacuum above. This is known as *Torricelli's Experiment*. He thus disproved the previous view that "Nature abhors a vacuum," and was led to infer that the pressure of the atmosphere on any area equals that of a column of mercury about 30 inches high and of a cross-section equal to the area. On hearing of Torricelli's experiment, Pascal reasoned that the pressure should be less and the column of mercury in Torricelli's tube lower at the top of a mountain and he wrote to a relative, who lived near the Puy de Dome in Auvergne, to make the test. The result confirmed his conjecture.

219. The Mercurial Barometer. Torricelli's tube was the first and simplest barometer or pressure-gauge for measurement of the pressure of the atmosphere. The most accurate mercurial barometer of the present day is a Torricellian tube with a scale and vernier for accurate measurement of the height of the mercury column, and a device by which the mercury in the cistern may be readily brought to a definite height. In Fortin's *cistern barometer* the cistern, C, has a flexible leather bottom, S, the center of which rests on a screw, V. By turning the screw the level of the mercury in the cistern can be raised or lowered so that when the barometer is read the level of the mercury in the cistern shall always be the same, namely, zero of the scale on which the height of the barometer is read. Without such an adjustment, the level of the mercury in the cistern would fall or rise as the height of the mercury in the tube, T, rose or fell. That the level of the mercury in the cistern may be observed the upper part of the cistern is of glass and a small ivory stud, O, projecting downward from the top of the cistern is adjusted by the maker so that its end is on a level with the zero of the scale.

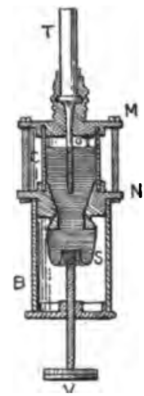


FIG. 118. Cistern of Fortin's barometer.

The image of the stud in the surface of the mercury is observed and when, as the level of the mercury is raised by the screw, the end of the stud and the end of its image just meet, the surface of the mercury is at the zero of the scale. In filling such a barometer care must be taken that no air remains in the mercury, and for this purpose after the tube has been filled it is inverted and the mercury boiled so that the air is expelled. The mercury in the cistern becomes somewhat tarnished in course of time and the image of the stud ceases to be distinct.

A simpler form of barometer is Bunsen's *siphon barometer*. In this there is no cistern, but the lower end of the tube is turned vertically upwards. The difference of level in the open and in the closed end is the barometric height. Thus readings of both ends of the mercury column are necessary. Scales are etched on

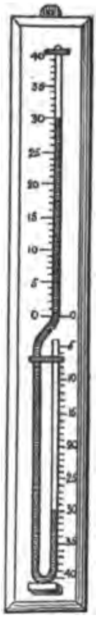


FIG. 119. Bunsen's siphon barometer.

both branches; the one on the longer arm reads upwards and that on the shorter arm reads downwards. The two scales are usually laid off with the same position for the zero, so that the sum of the two readings is the height of the barometer.

Another form of barometer is the *Aneroid* (Greek anēros = dry) barometer in which no liquid is used. It consists of a metallic box exhausted of air, with a thin metallic cover. Changes in atmospheric pressure cause slight changes of curvature in the cover, and by means of a multiplying system of levers these changes are transmitted to a pointer, which moves around a circular scale that is graduated in cms. or inches so as to correspond to the readings of the mercurial barometer. This form of barometer is more convenient for travellers, but it has the disadvantage that its index must frequently be reset by comparison with the mercury barometer.

220. Uses of the Barometer. A knowledge of barometric pressure is of great importance in weather forecasting. The governments of the United States and other civilized nations maintain a large number of stations where records of the barometer are kept. From simultaneous readings over a wide area the direction in which storms (or areas of low pressure) will move can be predicted. Such predictions lead annually to the saving of thousands of lives, and of much valuable property in shipping.

Since the atmospheric pressure is less at higher levels, it is possible to ascertain the height of a mountain by observing the atmospheric pressure at the top and at the bottom. Near sea-level the height of the barometer diminishes by about 0.1 inch for every 80 feet of ascent; but as the elevation increases the rate of fall diminishes owing to the greater rarity of the air. Allowance must be made for any difference of temperature at the two stations of observation.

221. Pressure and Volume of a Mass of Gas. Common observa-

tion shows that added pressure on a mass of gas diminishes its volume. Thus, in pumping up a bicycle tire, a large volume of air from the atmosphere is forced by high pressure into the small volume of the tire. Conversely, diminution of pressure allows a gas to expand. Against the pressure exerted on a gas it exerts an equal and opposite pressure, so that it is immaterial whether we speak of the pressure *on* or pressure *of* a gas.

The law connecting the volume and the pressure of a gas is extremely simple, but it was not discovered until 1662, the discoverer being Robert Boyle. (Fourteen years later Mariotte re-

FIG. 120. Boyle's tube for pressures greater than atmospheric.

FIG. 121. Boyle's tube for pressure less than atmospheric.

discovered the same law.) ***The volume of a gas at constant temperature varies inversely as its pressure,*** or, denoting the pressure and volume by p and v respectively, $pv =$ a constant. Boyle discovered this law by experiments conducted with a tube bent as in figure 120, the shorter arm being closed and containing air and mercury, while the longer was open and was filled to varying depths with mercury. If to the difference of level in the two arms the height of the mercury barometer at the time be added, the sum is proportional to the pressure on the air, while the length of the tube occupied by air is proportional to the volume of the air. Thus he discovered the truth of the law for pressures exceeding an atmosphere. For pressures below an

atmosphere he used a straight tube containing, initially, air and mercury and closed at one end; the open end was then plunged into a deep vessel of mercury. By drawing the tube to different heights the volume of the air increased with diminishing pressure. Thus Boyle verified the law for pressures less than an atmosphere.

222. Deviations from Boyle's Law. While the law stated by Boyle is accurate enough for all ordinary practical purposes, careful tests have shown that it is not perfectly accurate. The most complete tests were made by Amagat. He found that in the case of air, while the pressure is being increased from one atmosphere to about 78 atmospheres, pv steadily diminishes until its value is .98 of its value at one atmosphere. Thereafter, with increasing pressure, pv increases so that at 3000 atmospheres it has a value 4.2 times its initial value. In the first stage (that is up to 78 atmospheres) v decreases more rapidly than Boyle's law would indicate; thereafter it decreases less rapidly, so that at 3000 atmospheres its volume is 4.2 times what it would be if Boyle's law were perfectly accurate. (It may be noted that at 3000 atmospheres air has a density of .93, nearly equal to that of water; while the density of liquid oxygen at its critical pressure is about .7 and that of liquid nitrogen about .4.)

Other gases, excepting possibly hydrogen, show similar deviations from Boyle's law; but the pressure at which pv is a minimum is widely different for different gases, and so, too, is the magnitude of this minimum value of pv.

In his earlier experiments (1881) Amagat measured pressures by a very tall manometer in a mine shaft. Later he designed a special gauge for very high pressures. This consisted of two opposed pistons of very different diameters in separate cylinders. The high pressure, P, was applied to the small piston, of area a, and was counterbalanced by a much smaller pressure, p, applied to the large piston, of area A. Evidently $Pa = pA$, and, p being measured by a mercury manometer, P was deduced. Very viscous liquids were used in the cylinders to diminish leakage.

Starting with the view that a gas consists of flying particles the impact of which produces the pressure observed in a gas, Van der Waals deduced the following formula which agrees very well with the results of Amagat's experiments,

$$(p + a/v^2)(v - b) = \text{a constant},$$

at constant temperature, a and b being constants that are different for different gases.

223. Modulus of Elasticity of a Gas. The shear modulus of a gas being zero, a gas has only one modulus, namely the bulk modulus, and this is (when the gas is kept at constant temperature) simply equal to the pressure, p, of the gas. This is seen from Boyle's law. For when the pressure is p and the volume v, let an additional small pressure x be applied and let the volume be thereby reduced by the small quantity z, then by Boyle's Law

$$(p+x)(v-z)=pv;$$

or if we neglect the product of the small quantities x and z

$$vx=pz$$

Now the bulk modulus is the increase of pressure x divided by the proportional decrease of volume z/v, and from the last equation this is equal to p.

224. Buoyancy of a Gas. A body such as a balloon, lighter than the volume of air which it displaces, will ascend in the air when released. The force giving it an acceleration upwards equals the difference of its weight and the weight of the air which it displaces. If it rises to such a height that its mean density equals the density of the rarefied atmosphere, it will not ascend unless lightened by casting some of its load overboard. A large man displaces about ¼ lb. of air. When a body is weighed in air with weights that are supposed correct if used in a vacuum, the true weight of the body will not be obtained unless correction be made for the effect of the buoyancy of the air.

225. Manometers. A manometer is an apparatus for measuring the pressure of a fluid. In the simplest form the pressure to be measured is balanced against the pressure of a column of liquid in a tube. This is called the *open tube manometer* or siphon gauge. The pressure is found from the difference of level of the liquid in the two arms and the density, ρ, of the liquid. In absolute units of force $p=g\rho h$, while in the weight of unit mass as unit of force $p=\rho h$.

In another manometer the pressure to be measured is balanced against that of a gas (usually air) in a uniform *closed tube.* By

FIG. 122. Open tube manometer.

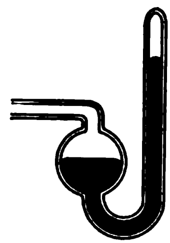


FIG. 123. Closed tube manometer.

Boyle's Law the pressure is inversely as the volume, that is inversely as the length of the air column.

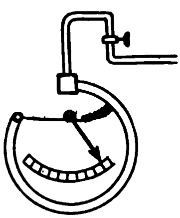


FIG. 124. Bourdon's pressure-gauge.

In *Bourdon's Pressure Gauge* a hollow tube of metal having an elliptical cross-section is bent into an arc of over 180°. One end of the tube is closed. When the fluid of which the pressure is to be measured is admitted to the open end, the curved tube will become less curved under increased pressure and more curved under decreased pressure. An index moving over a scale is attached to the free end. The action depends on the fact that the pressure tends to increase the interior volume of the tube; and, since a circular cross-section allows of more volume than an elliptical one for a given periphery, the section will under increased pressure tend to the circular form and the change of form of the cross-section causes the change of shape of the tube.

226. Viscosity of Gases. The viscosities of gases are small compared with those of liquids. Thus the viscosity of air is about

$1/80$ of that of water. While the viscosity of air is small, it is sufficient to retard greatly the fall of small particles of dust and small drops of water such as constitute a cloud. At the height of a cloud where the air is about one thousand times less dense than water, a drop of water one thousandth of an inch in diameter falls about 0.8 inch per second, while a drop one ten-thousandth of an inch in diameter falls about one hundred times more slowly or about 0.5 inch in a minute. For large drops such as constitute rain the viscosity of air offers practically no resistance; the resistance which prevents such drops attaining enormous velocities is the inertia of the air.

The viscosity of a gas increases when its temperature rises, which is the opposite of the case with liquids. The viscosity of a gas at constant temperature does not change appreciably when its density is altered by change of pressure.

227. The Kinetic Theory of Gases. The view that a gas consists of a myriad of particles in incessant motion may be regarded as firmly established. The evidence for this belief is that we can from it deduce nearly all the properties of a gas, and the agreement between these deductions and the observed facts could hardly be a mere accidental coincidence. As we do not yet know the details of the structure of the particles of which a gas consists, there are some properties of a gas which we cannot yet deduce from this theory. A single contradiction between the numerous known properties of a gas and the deductions made from the theory would be fatal to the latter; no such contradiction has ever been found.

As an illustration of the way in which the theory accounts for the properties of gases we shall show that it explains Boyle's Law.

Before doing so we must state the theory more in detail. The following, while an incomplete statement, will be sufficient for our purposes. (*a*) A single gas consists of particles all of the same size moving in random directions; (*b*) when the particles impinge on one another and on the walls of the vessel, they rebound like smooth spheres with a coefficient of restitution of unity; (*c*) unless a gas is greatly condensed, the particles are so far apart compared with their dimensions that they do not exert any appreciable force on one another except at impact. It will be noticed that we do not assume that the velocities of all the particles are the same,

and in fact there is good ground for believing that the velocities differ considerably.

For simplicity, consider a gas contained in a rectangular vessel the edges of which are a, b and c in length, and let A_1 and A_2, each of area bc, be perpendicular to the edges of length a. Let us first fix our attention on some particular particle which has a velocity V in some direction. V may be resolved into three components u, v, w, in the directions of the edges respectively, u being in the direction of the a sides. Suppose the particle to impinge on the side A_1. The force that it will exert on that side at impact will depend on its mass and on u, not at all on v and w. If it impinged without rebounding it would give momentum equal to its mass, m, multiplied by u, or mu. But it rebounds with a velocity the component of which perpendicular to A_1 is u in the opposite direction; hence the momentum it gives to A_1 is $2mu$. Let us now suppose for the present that it reaches A_2 without impinging on any other particle; for this it will require a/u seconds. At A_2 it will rebound with a velocity the component of which perpendicular to A_2 is u, and will, supposing it to encounter no other particle, reach A_1 in time $2a/u$ when it will again rebound. Hence in every second it will impinge $u/2a$ times on A_1, and in every second it will give to A_1 momentum $2mu \cdot u/2a$, or mu^2/a. The total force exerted on A_1, that is the momentum imparted to A_1 per second, is the sum of mu^2/a for all the particles, and to find the pressure p on A, we must divide this sum by the area of A_1, namely bc. Hence

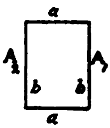


FIG. 125.

$$p = \frac{m}{abc}(u_1^2 + u_2^2 + \cdots)$$

Let us now denote the total number of particles in the vessel by N, and the number per unit volume by n. Since abc is the total volume of the vessel, $nabc = N$. Hence

$$p = mn\,\frac{(u_1^2 + u_2^2 + \cdots)}{N}$$

The product mn is the mass of all the particles in unit volume, that is the density ρ; and $\dfrac{u_1^2 + u_2^2 + \cdots}{N}$ is the average value of u^2 for all the N particles in the vessel. Denoting this by $\overline{u^2}$ we see that $p = \rho\overline{u^2}$. For any one particle

$$V^2 = u^2 + v^2 + w^2,$$

and, since the particles are moving wholly at random, the average values of u^2, v^2 and w^2 are all equal and the value of each is therefore $\frac{1}{3}$ of the average value of V^2 which we may denote by $\overline{V^2}$. Hence

$$p = \tfrac{1}{3}\rho \overline{V^2}.$$

If v be the volume of a mass M of gas, since $\rho = M/v$,

$$pv = \tfrac{1}{3}M\overline{V^2}$$

The total kinetic energy of translation of the gas is the sum of the kinetic energies of translation of all the particles and is evidently equal to $\tfrac{1}{2}M\overline{V^2}$ or $\tfrac{3}{2}pv$. Now there is good reason to believe (§ 320) that if the temperature of a gas is constant, this kinetic energy is constant. Hence the product of the pressure and volume of a gas at constant temperature is constant, and this is Boyle's Law.

In the above we have neglected the fact that a particle may, during its passages between A_1 and A_2, impinge on other particles. If such an impact take place between two particles moving along a line perpendicular to A_1 and A_2, the particles will exactly exchange their velocities (§ 174), since they are of the same mass; and the second particle will therefore have in the a-direction a component equal to that of the first particle before impact. Thus the second particle will take the place of the first in the process described above. When the immense number of particles and the random nature of their motions are considered, it is seen that the effect is the same as if all the impacts were in the directions of u, v, and w.

228. Occlusion and Surface Condensation. When a gas is in contact with a solid there are molecular forces drawing the particles together, and these produce more or less condensation of the gas on the surface of the solid. This makes it impossible to remove the last traces of a gas from a glass vessel by means of an air pump. It also accounts for the fact that when a figure is traced on a sheet of glass by a stick the figure will appear when the glass is breathed on. The breath condenses less readily on the part of the glass that has been freed from condensed gas by the scraping of the stick.

A porous solid is readily permeated by a gas and condensation on the surfaces of the pores takes place. This is called *occlusion*. Very porous wood-charcoal will absorb nine volumes of oxygen, thirty-five volumes of carbonic acid and ninety volumes of ammonia per volume of the charcoal, and cocoanut-charcoal will absorb still more. This is why charcoal is so useful as a deodorizer. Platinum in the porous form called platinum sponge will absorb 250 times its own volume of oxygen. Palladium will absorb more than one thousand volumes of hydrogen. Its own volume is thereby increased by about one-tenth. The hydrogen

is therefore reduced to one thousandth of its original volume; to produce such a condensation by pressure alone would require a pressure of several tons per square inch.

229. Diffusion of Gases. Gases on account of their greater mobility diffuse much more rapidly than liquids. When two vessels containing different gases are connected by a wide tube, diffusion proceeds with great rapidity and in a short time each gas is found distributed in both vessels as if the other gas were not present. If one of the gases be a colored gas, such as chlorine, the process of diffusion can be observed. As regards the final result each gas acts to the other as a vacuum, but in the process of diffusion each gas retards the other. Gravity also plays some part in the process though not in the final result. Thus if the gases be hydrogen and carbon dioxide, the final mixture is attained more rapidly when the carbonic acid is in the higher vessel.

In the process of diffusion of two gases into each other each gas follows the same law as holds for the diffusion of two liquids, that is, each gas diffuses from places where the concentration of that gas is great to places where it is less and the rate of diffusion is proportional to the rate of fall of concentration.

230. Effusion and Transpiration of Gases. When a gas passes through a small aperture in a very thin plate the process is called *effusion*. Different gases escape from such an aperture at very different rates. The rate of effusion may be deduced from the principle of energy. Each part of the gas as it escapes has a certain velocity and therefore a certain kinetic energy, and this must equal the work performed by the pressure in the vessel in forcing the gas out. Let p be the excess of the pressure in the vessel over the external pressure. During the escape of a small volume v of the gas the pressure p does the same amount of work as if it had pushed out a piston in a cylinder. Hence (§ 195) the work done is pv. If the density is ρ the mass of the volume v of the gas is $v\rho$, and if its velocity is V its kinetic energy is $\frac{1}{2}v\rho V^2$. Equating the work done to the kinetic energy which it produces, we get

$$V = \sqrt{\frac{2p}{\rho}}$$

Thus the rate of escape is directly as the square root of the pressure and inversely as the square root of the density.

Bunsen's method of comparing the densities of gases consists in comparing their rates of escape through the same aperture under the same pressure.

In establishing the above formula we have supposed that no work is done against internal friction such as there would be if the escape were through a tube. The wall of the vessel was supposed very thin so that the diameter of the opening might be larger than the thickness of the wall. Yet even in this case there is some slight viscous friction. This friction is different for different gases; hence the above simple formula does not give the ratio of the densities very accurately. When a mixed gas escapes by effusion the composition of the escaping gas is not altered as it escapes.

When a gas escapes through a tube of small bore the process is called *transpiration*. In this case there is a resistance due to the viscosity of the gas and the rate of escape, for small differences of pressure, is proportional to the difference of pressure, not as above to the square root of the difference of pressure. The composition of a mixed gas is not altered by transpiration.

When a gas escapes through a porous partition in which the pores are very small, such as fine unglazed pottery-ware, the circumstances are different from those of the above cases. The pores are comparable in size with the molecules of the gas and, as might be expected, the rates of escape of different gases are so different that the constituents of a mixed gas escape at different rates. This affords a method of partially separating the constituents of a mixed gas, and as the process may be repeated several times the separation may be made nearly complete. By this process it has also been possible to show that the molecules of a single gas are all of the same size, since no separation can be produced by the above method.

231. Passage of a Gas through Rubber. Some gases also escape through membranes such as rubber and wet parchment, in which there are no pores in the ordinary sense. The gas is dissolved by the membrane on one side and given up on the other side, so that the passage through the membrane is a diffusion from parts of the membrane where the concentration is

greater to parts where it is less. The same is true of the passage of a gas through a film of liquid. In a somewhat similar way hydrogen will pass through a red-hot platinum and iron.

232. Pumps for Liquids. The oldest form of pump, or *suction pump,* consists of a piston moving in a cylinder or barrel which is connected with the well by a pipe. In the pipe, or at the top of the pipe, there is a valve called the inlet valve which can open towards the cylinder but not in the opposite direction; and in the piston there is a valve called the outlet valve which can open outwards but not inwards towards the cylinder. When the

FIG. 126. Suction pump.

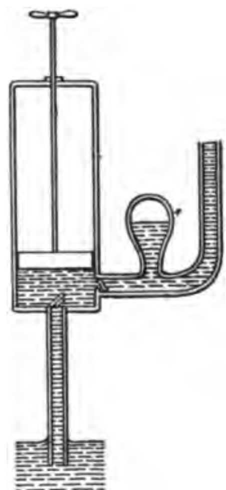


FIG. 127. Force pump.

piston is first raised the air in the cylinder expands and its pressure diminishes. The outlet valve closes owing to the excess of pressure on the outside, and, for the same reason, the inlet valve opens and air from the suction pipe enters the cylinder. Thus the air in the suction-pipe is rarefied and the greater atmospheric pressure on the water in the well forces water some distance up the suction-pipe. After some strokes the water enters the cylinder and flows out by the outlet valve.

Since it is the pressure of the atmosphere that raises the water in

a suction-pump, water cannot be raised by this means higher than atmospheric pressure will raise water in a vacuum; and since the density of water is to that of mercury as 1 to 13.6, it follows that the maximum theoretical height is 13.6 times the height of the mercury in a barometer or about 34 feet. The practical limit of suction-pumps is considerably less than this, owing to the presence of air in water and to the difficulty of making the contact between piston and pump air-tight. When water is to be raised to a greater height a *force-pump* is used. This differs from the suction-pump in the fact that the outlet valve is not in the piston but in a side tube connected to the cylinder near the inlet valve. During each downward stroke of the piston water is forced up this side tube, and the height that may be reached will depend on the force that can be applied to the piston and the maximum pressure that the pump will stand without breakage of some part.

The outflow from the delivery tube of a force-pump as described above would be intermittent; but it may be rendered more nearly continuous by means of an "air chamber," in which air, being put under pressure by the water forced in, exerts continuous pressure on the outflowing water (Fig. 128).

233. The Siphon. The siphon is a bent tube for removing liquid from a vessel. The tube is filled with liquid and is then inverted and one end A is immersed in the liquid while the other end C is kept closed. When C is opened liquid flows through the tube and out through C so long as the C is below the level D of the surface of the liquid.

To explain the action of the siphon let us consider the pressure on the liquid at C before the end C is opened. If the difference of level of D and C is h, the pressure on the liquid at C is greater than atmospheric pressure by $g\rho h$. Hence, when C is opened, the excess of pressure inside causes a flow and the flow continues so long as C is below the level of D and A remains immersed. A siphon will not act if the highest point B of the tube is at a greater height above the level of D than the height to which atmospheric pressure will force the liquid in an exhausted tube.

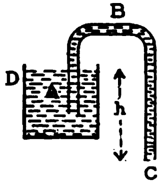


FIG. 128. Siphon.

234. Air-pumps. The first pump for removing air from a vessel was invented by Otto von Guericke (in 1650). It was essentially a suction pump like that used for water, the only difference being the closer fit of piston required to prevent leakage in the case of a gas. The degree of exhaustion that can be attained by such a pump is low. The flap-valve, at the end of the suction-tube, will not act automatically when the pressure in the receiver has become very small. For this reason a conical plug, carried by a rod that passed with some friction through the piston, was substituted. Another difficulty is caused by the fact that the piston cannot be made to fit the lower end of the cylinder with perfect accuracy so as to expel all the air drawn from the receiver into the cylinder. The latter defect has been remedied in the Geryk pump in which there is a layer of oil at the bottom of the cylinder; oil above the piston also prevents leakage at the piston valve.

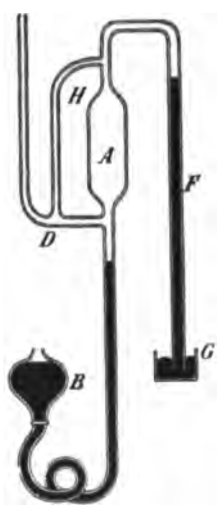


FIG. 129. Toepler's mercury pump.

235. Mercury Pumps. When a very high vacuum is needed glass pumps are used in which mercury plays a part somewhat analogous to that of a piston. Of these there are several forms.

In Toepler's form (Fig. 129) a glass bulb A is connected by a flexible tube with a mercury reservoir B which can be raised and lowered. A tube D connects the bottom of A with the vessel to be exhausted, while a long tube F is joined to the top of A and dips into a vessel of mercury G. As B is raised, mercury begins to flow into A, sealing the connection of A with D; as B is raised higher, mercury flows into A and D, expelling the gas through the mercury in G. When B is lowered, mercury rises in F and prevents the return of air through F; the connection between D and A is unsealed and gas flows from the vessel to be exhausted into A. This gas is expelled when B is again raised, and so on. The side tube H prevents the sudden inrush of gas into the bottom of A as B is lowered and thus saves A from danger of breakage.

By various mechanical devices the labor of raising and lowering B may be reduced or eliminated.

A new form of mercury pump of very high efficiency was invented by Gaede in 1907. Its principle is indicated (without details) in Fig. 130 and Fig. 131. An iron cylinder, g, with a glass

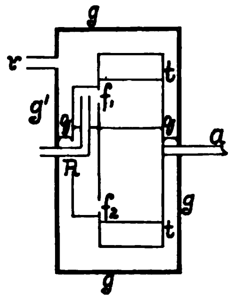


FIG. 130.

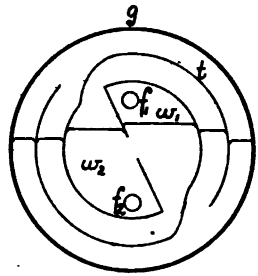


FIG. 131.

face, g', is more than half filled with mercury, the surface of which is at q. Inside of g there is a porcelain drum, t, rotating about an axis, a, which passes air-tight through g. This drum is divided into two chambers, w_1 and w_2, which communicate with g by long channels between the division-walls of t. Each chamber has an opening, f, by which the part of the chamber above the mercury is connected, through the tube R, with the receiver to be exhausted. As the drum is rotated counter-clockwise, the chamber w_1 is gradually emptied of mercury and filled with air drawn in through R. As the rotation continues, f_1 is immersed, and the air in w_1 is driven into g. The action of w_2 is similar. Since either f_1 or f_2 is always out of the mercury, the suction through R is continuous. The air in g is removed by another pump (which may be much less efficient) connected to r. Gaede's pump will produce a vacuum of about 0.00004 mm.

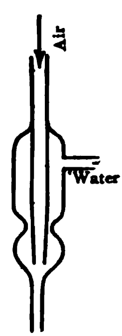


FIG. 132. Bunsen's aspirator.

236. Bunsen's Aspirator is a form of pump adapted for use on a water supply. A tube connected with the vessel to be exhausted is sealed into an elongated bulb. The water is forced into the bulb and, in escaping through the lower open end of the bulb, it drags

the air out of the tube along with it. With sufficient water pressure a vacuum of about two centimeters (mercury gauge) may be obtained by means of this aspirator.

REFERENCES.

CREW'S *Principles of Mechanics* contains a brief and very clear account of the subject stated in elementary Vector and Calculus language.

POYNTING & THOMSON'S *Properties of Matter* is especially valuable for information on gravitation, elasticity and properties of fluids.

The above-mentioned books will be found useful for somewhat advanced systematic study.

MACH'S *Principles of Mechanics* is a very interesting and elementary account of the historical development of the subject.

COX'S *Mechanics* is an elementary book with notes on the historical development.

WORTHINGTON'S *Dynamics of Rotation* is an elementary book with numerous suggestive experiments.

PERRY'S *Spinning Tops* is a popular account of the principles of the gyroscope.

LOVE'S *Theoretical Mechanics* is a very careful account of the logical relations of the parts of the subject.

MAXWELL'S *Matter and Motion,* while elementary and very brief, is a masterpiece by a great physicist.

TAIT'S *Properties of Matter* contains an elementary treatment of gravitation, elasticity and properties of fluids.

LODGE'S *Pioneers of Science* consists of popular lectures on Galileo, Newton, etc.

POYNTING'S *Mean Density of the Earth* describes all the methods used.

Encyclopedia Britannica, articles on "Weights and Measures," "Mechanics" (Tait), "Constitution of Bodies" (Maxwell), "Elasticity" (Kelvin).

DANIELL'S *Principles of Physics* is a large compendium.

PROBLEMS.

Velocity and Acceleration.

1. A train acquires 5 minutes after starting a velocity of 40 km. per hour. Assuming constant acceleration, what is the distance passed over during the 5th minute? Ans. 0.6 km.

2. A train having a speed of 70 km. per hour is brought to rest by brakes in a distance of 600 m. What is the acceleration (assumed constant)? Ans. — 408.3 km./min².

3. What is the final speed of a body which, moving with uniform acceleration travels 72 meters in 2 minutes if:

(a) the initial speed = 0?
(b) the initial speed = 15 cm. per sec.?

Ans. 120 cm./sec.; 90 cm./sec.

4. A body is projected at an angle of 30° with the horizon with a velocity of 30 m. per sec. When and where will it again meet this horizontal plane? How far will it ascend? Ans. 3.06 s.; 19.5 m. 11.4 m.

5. A body slides down an inclined plane and in the 3d sec. describes 110 cm. What is the inclination? Ans. $2° 35'$.

6. What initial vertical velocity must a ball have in order to fall back to its starting-point in 10 sec.? Ans. 4900 cm./sec.

7. At what angle with the shore must a boat be directed in order to reach a point on the other shore directly opposite, if the speed of the boat be 4 miles per hour and that of the stream be 2 miles per hour? Ans. 60°.

8. A point goes over a circular path 10 cm. in diameter 4 times a second, at a uniform speed. To what acceleration is it subject? Ans. 6317 cm./sec^2.

9. A ball rises to a height of 50 ft. and travels 200 ft. horizontally. Find the direction and magnitude of the velocity with which it is thrown. Ans. $\theta = 45°$; $v = 82.2$ ft./sec.

10. Show that the time of descent (without friction) down all chords of a vertical circle is the same.

11. What velocity must a boy give a sling of 80 cm. radius in order that the stone shall not fall out of the sling? Ans. 280 cm./sec.

Force and Mass.

12. What force will a man who weighs 70 kg. exert upon the floor of an elevator descending with an acceleration of 100 cm. per sec. per sec. If ascending with the same acceleration? Ans. 77.1; 62.8 kg. wt.

13. A force of 1000 dynes acts upon a mass of 1 kg. for 1 min. Find the velocity acquired and the space passed over in this time. Ans. 60 cm./sec.; 1800 cm.

14. A shot weighing 10 lbs. is shot from a gun weighing 3 tons with an initial velocity of 1200 feet per sec. What is the initial velocity of the recoil? Ans. 2 ft./sec.

15. Three forces, 5, 12, 15 are in equilibrium. Find the angles between them. Ans. $62° 11'$; $134° 58'$; $162° 51'$.

16. A cord passes over two fixed pulleys and through a third pulley suspended between them. A mass of 10 g. is attached to one end of the cord, a mass of 5 g. to the other end, and the suspended pulley and attached weight weigh 2 g. The cords being all vertical, what are the accelerations of the three masses? Ans. 809 cm./sec^2.; 639; 724.

17. Twelve bullets are divided between two scale pans connected by a cord passing over a pulley. What distribution will produce the greatest tension on the support of the pulley?

18. Bodies of mass 10 kg. and 8 kg. are connected by a string over a pulley. How far does each move from rest in the first two seconds? Ans. 218 cm.

19. A baseball whose mass is 300 g. when moving with a velocity of 20 m. per sec. is squarely struck by a bat and then has a velocity of 30 m. per sec. in the other direction. Calculate the impulse and average force if the contact last .02 sec. Ans. 15000 g.—cm./sec.; 7.5×10^5 dynes.

Work and Energy.

20. With how much energy must a bullet weighing 20 g. be shot horizontally from a gun 4 m. above a level plane, in order to strike the ground 300 m. away from the gun? Ans. 1.11×10^{10} ergs.

21. A projectile traveling at the rate of 700 ft. per sec. penetrates to the depth of 2 in. Find the velocity necessary to penetrate 3 in. Ans. 857 ft./sec.

22. A hammer weighing 6 kg. and moving with a velocity of 100 cm. per sec. drives a nail into a plank 1 cm. What resistance does it overcome (supposed uniform)? Ans. 3×10^7 dynes.

23. A man can bicycle 12 miles an hour on a smooth road; downward force of each foot in turn = 20 lbs., length of stroke = 1 ft., bicycle is advanced 17 ft. for each revolution of the cranks. At what H. P. does he work? Ans. 1.24 H. P.

24. A man, weight 180 lbs., runs up 26 steps, each 7 in. high, in 4 sec. At what H. P. does he work? Ans. 1.2 H. P.

25. A sprinter who weighs 161 lbs. runs 40 yds. in $4\frac{2}{5}$ sec., 60 yds. in $6\frac{2}{5}$ sec., 100 yds. in 10 sec. What is (a) his velocity, (b) his kinetic energy, at the end of the 40 yards. (c) Calculate the rate of working in H. P. required to produce this kinetic energy (d) In what other ways does he expend energy. Ans. (a) $33\frac{1}{3}$ ft./sec. (b) 2777 ft. lbs. (c) 1.09 H. P.

26. Find the number of watts in one horse-power. Ans. 746.

27. A sprinter does 100 yards on the horizontal in 10.5 sec., and the same distance up hill with a rise of 32 ft. in 17.5 sec. Assuming that his rate of working is the same throughout, calculate the added work done in the additional 7.5 seconds up hill and the rate of working that this implies. Ans. 1.33 H. P.

28. 100 cu. ft. of water pass over a dam 10 ft. high in 1 min. What horse-power could be derived from this if all were utilized? Ans. 1.9 H. P.

29. A 30-gram rifle bullet is fired into a suspended block of wood weighing 15 kilos. If the block is suspended by a string of length 2 meters and is moved through an angle of 20°, calculate the velocity of the bullet. Notice that the impact of the bullet on the block does not change the total *momentum* of both (§ 46) and during the subsequent swing of the pendulum its total *energy* remains constant. Ans. 770 m./sec.

Rotation.

30. If a locomotive driving-wheel 1.5 m. in diameter makes 250 revolutions per minute, what is the mean linear speed of a point on the periphery? Of the point when it is highest? When it is lowest? Ans. 19.8 m./sec.; 39.6 m./sec.; 0.

31. The armature of a motor revolving at the rate of 1800 revolutions per minute comes to rest in 20 seconds after the current is shut off. Calculate its average angular acceleration and the number of revolutions. Ans. —9.42 red./sec.; 300 rev.

32. Find in radians per second the angular velocity of the earth about its axis and deduce the component of this angular velocity about a diameter through a point in latitude 40° (Principle of Foucault's pendulum).

Center of Mass.

33. A circle has a diameter of 16 cm. a smaller circle tangent to it and 12 cm. in diameter is cut out of it. Where is the center of gravity of the remainder? Ans. 10.6 cm. from common tangent.

34. Two cylinders of equal length (= 20 in.), and having diameters of 12 and 6 in., are joined so that their axes coincide. Where is the center of gravity? Ans. 6 in. from junction.

35. Find the center of gravity of a table 4 ft. × 3 ft. × 1 in., with legs at the corners 2 ft. × 2 in × 2 in. Ans. 2.7 ft. from top.

36. The mass of the moon is $\frac{1}{80}$ of that of the earth and the average distance between their centers is 240,000 miles. Calculate the position of the center of mass of the two. Ans. 2963 m. from center of earth.

37. At the corners of an equilateral triangle ABC masses of 1, 2 and 3 lbs. respectively are placed. Find the distance of their center of mass from BC assuming each side of the triangle to be 1 ft. in length. Ans. 0.144 in.

Moments.

38. A bar 6 ft. long and pivoted at the middle has a weight of 24 lbs. hung at one extremity. What is the moment of the weight (a) when the bar is horizontal, (b) when it makes an angle of 30° below, and (c) of 60° above with the horizontal position? Ans. 72; 62.3; 36 lbs. wt. ft.

39. If it is wished to upset a tall column by a rope of given length pulled from the ground, where should it be applied?

40. Find the moment of inertia of a sphere ($m = 20$, $r = 2$) about an axis tangent to its surface. Ans. 112.

41. Find the moment of inertia of three circular disks tangent to each other in the same about a perpendicular axis passing through the center of one of them. The mass of each is 100 g. and the radius of each is 6 cm. Ans. 17,100 gm. cm^2.

42. Two masses, 100 kg. and 200 kg., respectively, are connected by a rigid rod 1 m. long. The system is thrown so that the center of gravity has a velocity of 20 m. per second and the system turns 10 times per second about this center. Find the kinetic energy of the system.
Ans. 192×10^{10} ergs.

43. What energy has a grindstone $1\frac{1}{2}$ m. in diameter, weighing 1000 kg. and rotating once every 2 sec.? Ans. 13.9×10^{8} ergs.

44. A solid iron cylinder, 100 cm. diameter, rolls down a plane 6 m. long inclined at 30°. What linear velocity does it acquire?
Ans. 627 cm./sec.

45. A block of stone weighs 2.5 tons and is in the form of a cube of 1 yard side. It rests on level ground. What is the least force which applied to the block will cause it to revolve about a horizontal edge?
Ans. 1768 lbs. wt.

Parallel forces.

46. Parallel forces of 1 2 and 3 units respectively act at the corners *A, B, C* of an equilateral triangle of 1 ft. side. Find the distance of the resultant from *BC*. Ans. 0.144 in.

47. Parallel forces of 10 and 6, but in opposite directions, are applied to a bar at distances of 8 and 3 from one end. What is the magnitude of the resultant and where does it act? Ans. 4; 15.5.

48. Two equal parallel forces, each 50 dynes, act in opposite directions at the ends of a bar 10 cm. long. The bar makes an angle of 45° with the direction of the force. What is the moment of the couple?
Ans. 353.5 dynes-cm^2.

49. A man and a boy carry a weight of 20 kg. between them by means of a uniform pole 2 m. long, weighing 5 kg. Where must the weight be placed so that the man may carry twice as much of the whole weight as the boy? Ans. 0.416 m. from middle.

Equilibrium.

50. A rod, the mass of which is 1 kg., hangs from a hinge on a vertical wall and rests on a smooth floor. Calculate the force on the floor and the force on the hinge. Ans. 500 g.; 500 g.

51. A uniform ladder 30 feet long and of 50 lbs. weight rests with the upper end against a smooth vertical wall, and the lower end is prevented from slipping by a peg. If the inclination of the ladder to the horizontal is 30°, find the force on the wall and at the peg. Ans. 50; 66.1 lbs. wt.

52. A barn door is 10 ft. long and 5 ft. wide and weighs 200 lbs. The hinges are 1 ft. from the ends and the weight is carried entirely by the upper hinge. Find the direction and magnitude of the resultant force on the upper hinge. Ans. 524 lbs.; 17° 21′ to vertical.

Perodic motions.

53. One end of a certain rod is clamped. If the other end is pulled 1 cm. from its natural position and then released, it starts with an acceleration of 10 cm. per sec per sec. What is the period of its vibration? Ans. 1.98 sec.

54. The balance-wheel of a watch makes 5 complete vibrations in 2 sec. With what angular acceleration will it start when turned 30° from its position of equilibrium and released? Ans. 129.34 rad./sec^2.

55. A hoop of 25 cm. radius hangs on a peg. Calculate its period of vibration. Ans. 0.16 sec.

56. A clock gains 3 min. a day. Find the error in the length of the pendulum ($g = 980$). Ans. 0.414 cm.

57. A pendulum which is a seconds pendulum where $g = 980$, vibrates but 59.95 times a minute on top of a mountain. What is the acceleration of gravity at this point? Ans. 978.37.

53. A rod 2 m. long is freely suspended at one end. Calculate its period of vibration. Ans. 4.63 sec.

59. A second's pendulum is drawn aside and released and at the same moment a ball is allowed to fall. The ball and the bob collide as the pendulum passes through the vertical. Calculate the height of fall of the ball. Ans. 122.5 cm.

Friction.

60. The coefficient of friction for two surfaces $= 0.14$. A pull of 20 kg. weight will overcome what pressure between them? Ans. 143 kg.

61. What force applied parallel to a plane inclined at 20° will push up a block weighing 100 kg., the coefficient of friction between the two being 0.24: (*a*) the block moving uniformly; (*b*) the block having an acceleration of 100 cm. per sec. per sec.?

Ans. (*a*) 56.7 kg. wt.; (*b*) 66.9 kg. wt.

62. What is the coefficient of friction between a body and a horizontal plane if the body loses a velocity of 100 ft. per sec. and comes to rest in moving 200 ft. over the plane? Ans. 0.776.

63. A toboggan slides 100 yards down a track inclined at 20° to the horizontal in 11 seconds. Calculate the coefficient of friction. Ans. 0.20.

64. A small block rests on a horizontal revolving platform at a distance of 40 cm. from the axis of revolution. If the coefficient of friction is .30 at what angular velocity of the platform will the block just begin to slip. Ans. 1.71 rad./sec.

Machines.

65. A man raises a stone 1 in. with a lever of the first class 10 ft. long weighing 50 lbs., the fulcrum being 1 ft. from the point of application to the stone. If he exerts a force of 100 lbs. wt. what force is applied to the stone and what work does he do?

Ans. 1,100 lbs. wt.; 75 ft. lbs.

66. A boy who exerts a push of 50 lbs. wt. wishes to roll a barrel weighing 200 lbs. into a wagon $2\frac{1}{2}$ ft. high. Assuming that he pushes in a line through the center of the barrel parallel to the plank, how long a plank will he need and how much work will he do? Ans. 10 ft.; 500 ft. lbs.

67. A body weighs 12 lbs. on one side of a false balance and 12.5 lbs. on the other side. What is the ratio of the arms of the balance?

Ans. 12.247 lbs.

68. A man weighing 150 lbs. sits on a platform suspended from a movable pulley and raises himself by a rope passing over a fixed pulley. Supposing the cords are parallel, what force does he exert? Ans. 75 lbs. wt.

69. A wheel whose radius is 25 cm. is fastened to one end of a screw whose pitch is 1 mm. What force can the screw exert in its nut when a

force of 1 kg. wt. is applied tangentially to the wheel, friction being supposed negligible? Ans. 785 kg. wt.

70. Compare the mechanical advantages of a block and tackle when the end of the cord is attached to the upper block and when it is attached to the lower.

71. How far above the surface of the earth must a body be to lose 0.1 per cent in weight? Ans. 1.95 mi.

Gravitation.

72. If the moon's mass is $\frac{1}{80}$ that of the earth, and its diameter 2160 miles, that of the earth being 7900 miles, what is the acceleration of gravity on the moon's surface? Ans. 164cm./sec^2.

73. Find the time of revolution of the earth which would cause bodies to have no apparent weight at the equator. Ans 1.41 hr.

74. A wire 300 cm. long and 1 mm. in diameter is stretched 1 mm. by a weight of 3,000 g. What is Young's Modulus? Ans 11.2×10^{11}.

Elasticity.

75. A weight is hung from the ceiling by a steel wire 2 m. long and of 1 mm. diameter joined to a copper wire 1 m. long and of 0.5 mm. diameter. Another weight sufficient to produce a total extension of 1 mm. is added. Calculate the extension of each part. Ans. 0.19 mm.; 0.81 mm.

76. To opposite faces of a cubical block of jelly of 20 cm. edge parallel and opposite forces of kg. each are applied and produce a relative motion of 1 cm. Calculate the strain, the stress and the shear modulus.
Ans. 0.05; 2450 dynes/cm.^{2}; 49000.

77. An iron bar of 400 c.c. volume falls from a ship and sinks to the bottom of an ocean 1000 m. deep. How much is its volume diminished, assuming that each 10 m. of water pressure produces a pressure equal to that of the atmosphere, which equals one million dynes per sq. cm.
Ans. 0.024 c.c.

78. A ball weighing 20 kg., moving with a velocity of 500 cm. per sec., strikes a second ball weighing 100 kg. which is at rest. If the first ball rebounds with a velocity of 100 cm. per sec., what will be the velocity of the second? Ans. 120 cm./sec.

79. Two bodies differing in bulk weigh the same in water; compare the weights in mercury; in vacuo.

Properties of Liquids.

80. A mass of copper suspected of being hollow weighs 523 g. in air and 447.5 g. in water. What is the volume of the cavity. Ans. 16.8 c.c.

81. The specific gravity of ice is 0.918, that of sea-water 1.03. What is the total volume of an iceberg of which 700 cu. yds. is exposed?
Ans. 6348 cub. yds.

82. A block of wood weighing 1 kg., whose density is 0.7, is to be loaded with lead so as to float with 0.9 of its volume immersed. What weight of lead is required (1) if the lead is on top? (2) if the lead is below?
Ans. 286 g.; 313.5 g.

83. A hydrometer sinks to a certain mark in a liquid of sp. gr. 0.6, but it takes 120 g. to sink it to the same mark in water. What is the weight of the hydrometer? Ans. 180 g.

84. One of the limbs of a U-shaped glass tube contains mercury to the height of 0.175 m.; the other contains a different liquid to a height of 0.42 m., the two columns being in equilibrium. Required, the density of the second liquid with reference to mercury and to water.

85. Find the volume in cub. ft. of the smallest block of ice which, floating on fresh water, will just carry a man who weighs 150 lbs.
Ans. 29.3 cub. ft.

86. Given a body A which weighs 7.55 g. in air, 5.17 g. in water, and 6.35 g. in another liquid *B*; required, the density of the body *A* and the liquid *B*. Ans. 3.17; 0.504.

87. A block of brass 10 cm. thick floats on mercury. How much of its volume is above the surface, and how many cm. of water must be poured above the mercury so as to reach the top of the block? (Density of mercury = 13.6; of brass = 8.5.) Ans. 0.375; 4.05.

88. Two tubes are inserted in a vessel of water on the same horizontal plane. The diameter of the one is 0.5 mm. and its length is 20 cm.; the diameter of the other is 0.25 mm. and its length is 10 cm. Compare the amounts of water flowing through the two tubes in a given time. Ans. 8 : 1.

89. The diameter of the small piston of an hydrostatic press is 2 in., the diameter of the large piston is 2 ft. What weight on the small piston will support two tons on the large piston? Ans. 27.77 lbs.

90. The pressure at the bottom of a lake is three times that at a depth of 2 m. What is the depth of the lake? (Atmospheric pressure = 76 cm. of mercury.) Ans. 7.52 m.

91. A retaining wall 3 m. wide and 40 m. long is inclined at 45° to the horizontal. Find the total force in kg. against it when the water rises to the top. Ans. 1.27×10^5 kg.

92. What is the outward force on the sides of a circular tank 1 m. in diameter, the height of the water being 150 cm.? What is the pressure on the bottom? Ans. 3536 kg. wt.; 1178 kg. wt.

93. The surface tension of a soap-bubble solution is 27.45 (dynes/cm.). How much greater is the pressure inside a soap-bubble of 3 cm. radius than in the air outside? Ans. 36.6 dynes.

94. How far will water be projected horizontally from an aperture 3 m. below the water level of a tank and 10 m. above the ground (neglecting air resistance)? Ans. 10.96 m.

Properties of Gases.

95. A body whose density is 2 is weighed in air of density 0.0013 with weights of density 9. The weight in air being 100 g., what is the true weight? Ans. 100.050 g.

96. If the barometer sinks 15 mm., how much is the pressure in dynes per sq. cm. decreased?
Ans. 19992 dynes/cm².

97. An air bubble at the bottom of a pond 6 m. deep has a volume of 0.1 c.c. Find the volume just as it reaches the surface, the barometer standing 760 mm. Ans. 0.0158 c.c.

98. Owing to the presence of air the mercury column in a barometer 85 cm. long stands at 70 cm. when an accurate barometer stands at 75 cm. What pressure will this barometer indicate when an accurate barometer stands at 72 cm.? Ans. 67.67 cm.

99. A barometer reads 73 cm. Calculate the thrust on one side of a board 1 m. square. Ans. 9928 kg. wt.

100. A barometer has a cross section of 2 sq. cm. and is so long that as the mercury stands at 76 cm., there is a vacuum space of 10 cm. long. Some air is allowed to enter and the mercury falls 10 cm. What was the volume of the air before it entered? Ans. 5.26 cm³.

101. How high must we ascend above the sea-level to observe a depression of 1 mm. in the height of the barometer? Density of air = 0.0013 (approx). Ans. 10.4 m.

102. A glass tube 60 cm. long, closed at one end, is sunk, open end down, to the bottom of the ocean. When drawn up it is found that the water has penetrated to within 5 cm. of the top. Atmospheric pressure = 76 cm. of mercury. Calculate the depth of the ocean, assuming the density constant. (Principle of Lord Kelvin's sounding apparatus.) Ans. 110.8 m.

103. In a vessel of 1 cu. meter volume are placed the following amounts of gas: (1) hydrogen, which occupies 1 cu. m. at atmospherc pressure. (2) nitrogen, which occupies 3 cu. m. at a pressure of 2 atmospheres. (3) oxygen, which occupies 2 cu. m. at a pressure of 3 atmospheres. Calculate pressure of mixture. Ans. 13 at.

104. The mouth of a vertical cylinder 18 in. high is closed by a piston whose area is 6 sq. in. If a weight of 100 lbs. be placed on the piston, how far will it descend, supposing the atmospheric pressure to be 14 lbs. per sq. in. and the friction negligible? Ans. 9.8 in.

105. A cylindrical diving-bell 7 ft. in height is lowered until the top of the bell is 20 ft. below the surface of the water. If the barometer height at the time is 30 in., how high will the water rise in the bell? What air pressure in the bell would just keep the water out? Ans. 2.96 ft.; 1.82 at.

HEAT.

By K. E. Guthe, Ph.D.

Professor of Physics in the University of Michigan.

INTRODUCTION.

237. Temperature. Experience makes us acquainted through the sensations of warmth and cold with a general physical property of bodies which is independent of their motions of translation or rotation and of any mechanical forces acting upon them. We can distinguish different degrees or qualities of this sensation and speak therefore of hot, warm, tepid, cool and cold bodies to characterize their effect upon our senses. The scientific term "temperature" is used to include all these physical states and to suggest a quantitative relation between them, by saying, for example, that a hot body is at a higher temperature than a cold one. *Temperature can therefore be defined as the physical measure of that condition of a body which enables it to produce the sensations of warmth and cold.*

By touching two objects we draw a conclusion as to their relative temperatures and if both bodies are made of the same material, may form in this way an approximately correct but rough estimate. Our conclusions become, however, entirely untrustworthy if we have to do with bodies of different material. A cold piece of iron will always appear cooler to the touch than a piece of wood of the same temperature. This is due to the fact that other properties of bodies besides their temperature—for example, conductivity for heat—greatly influence our estimate.

Thus while sensation furnishes us with the fundamental concept of temperature, we are forced, if we wish to assign a quite definite meaning to the word temperature to look for characteristic physical properties which change continuously and in a definite manner with temperature and upon which we can base a logical system of temperature measurements. (See Thermometry, § 240.)

238. Heat a Form of Energy. Any mass of a given substance at one temperature is physically different from what it is at any other temperature. A physical quantity which we call *heat* has been added to it or subtracted from it. A body may be heated in many different ways, as a kettle by direct contact with a fire, rooms by the warm air from the furnace, a nail by the blows from a hammer, a saw by friction against wood, an electric lamp by electricity, our bodies by the sun's rays and so forth.

Formerly heat was supposed to be a subtle, elastic fluid which upon entering a body would raise its temperature. This hypothetical fluid was called *caloric,* and its particles were assumed to repel each other, at the same time being attracted in different degrees by the various forms of matter. Since, after many controversies, it has also been demonstrated that the weight of a body does not change with its temperature, caloric must be without weight or some kind of imponderable matter. Many heat phenomena could be explained satisfactorily by this theory, which was generally accepted until the beginning of the nineteenth century; it failed however, when careful investigations at that time proved that excessive amounts of heat can be produced by friction alone.

The famous experiments of Count Rumford performed 1798 at the arsenal in Munich proved that practically unlimited amounts of heat can be produced by boring a cannon by means of a blunt drill, while the quantity of abraded metal was very small indeed. No stage was reached in the process, at which the heat, if it were a part of the metal, seemed exhausted. All that was necessary for a further development of heat, was the continued turning of the drill, *i. e.,* the overcoming of the friction between the drill and the metal, or the expenditure of mechanical work or energy.

In another way Sir Humphrey Davy showed in 1799 that heat can be produced by overcoming resistance due to friction. It was well known at that time that heat is needed to melt ice. By rubbing together two pieces of ice, the ice was melted, the resulting water being at the same temperature as the melting ice. In this case also it was apparently the energy spent in overcoming the friction between the two pieces of ice, which ultimately melted them.

.When a hammer strikes a nail the kinetic energy of the hammer suddenly disappears and heat appears. Innumerable examples of the same nature might be cited.

From analogy we may draw the conclusion, and correctly, as we shall see, that the heating of an electric light is accompanied by the disappearance of electrical energy, or that the warmth we feel in the sun's rays is due to a transformation of their energy into heat in our bodies.

Heat is therefore a form of energy considered mainly from the point of view of temperature changes produced by it and is measured by the temperature changes produced in a given mass of some standard substance, water. Quantity of heat is heat expressed in such arbitrarily chosen units (see § 260).

But we find almost all physical properties of matter influenced by heat, so that in general variations in the elastic, optical and electrical properties accompany temperature changes.

The temperature of a mixture of ice and water remains constant as long as both constituents are present. If the surrounding objects are warmer than the mixture, heat passes into the mixture, the surroundings are cooled, and the ice melts.

Formerly a distinction was made between "sensible" heat and internal or "latent" heat, which latter does not produce an effect upon the temperature. With the modern development of our interpretation of heat phenomena such a distinction has become unnecessary.

239. Molecular Theory of Heat. According to the molecular theory (§ 227), the smallest particles or molecules of a body are in a state of perpetual agitation. The constant collisions between the different particles must in the end produce a steady condition in which motion in one direction is as likely to occur as in any other. After such uniformly irregular motion has been established throughout, equal finite volumes of a homogeneous body contain the same amount of molecular kinetic energy.

Now let heat be added to a portion of such a body and let the body be left entirely to itself. According to the molecular theory of heat the heat energy serves, at least in part, to increase the kinetic energy of the molecules. For a short time the mean energy of motion of the molecules at the heated place will be larger than in the rest of the body, but it will soon distribute itself through the whole mass (conduction of heat) and in the

end there is again an equilibrium of uniformly irregular motion in which the average energy of each molecule is larger than before, and as a rule the temperature has become higher and again uniform throughout. Similarly two bodies, at different temperatures and in contact with each other, will come to the same temperature and be in thermal equilibrium; after which no further transfer of heat between them can be observed.

The heat necessary to melt a solid must also produce an increase in the energy of the molecules, in part probably as potential energy.

This theory explains heat as being molecular energy, partly kinetic and partly potential. The uniformly irregular motion of the molecules is the most natural one to be expected. Any regular motion of a system of molecules, such as those studied in Mechanics and other chapters of Physics, has the tendency, when disturbed, to become irregular molecular motion, as shown by the examples under § 238. On the other hand, heat can only be changed into specialized forms of energy under certain conditions, and on account of this difficulty of transformation and the large amount of heat which remains unavailable for practical purposes, it has unjustly been called the lowest form of energy. From this point of view we may speak of the energy of wind and waves being frittered down into heat or being dissipated as heat.

THERMOMETRY.

240. The Standard Thermometer. *Differences of Temperature* may be defined by any one of the numerous changes of physical properties which accompany a variation of temperature. An instrument designed for such purpose is called a *thermometer,* or if the temperature is high, a *pyrometer.*

Since the various properties of a body, or the same property of different bodies, are affected quite unequally by temperature changes, great differences in thermometric readings would be found, according to our choice of substance and particular property. For practical reasons it is, therefore, necessary that a specific property of a definite substance, upon which the construction of a standard thermometer can be based, should be selected arbitrarily. By comparison with such a standard all other thermometers can be calibrated.

We find for example that, when a gas is enclosed in a vessel of constant volume and heated, the pressure which it exerts upon the walls is increased. Now it has been agreed to choose the effect of temperature upon the pressure of hydrogen gas when inclosed in a vessel of constant volume as the principle upon which to construct a standard thermometer and *a change in temperature is defined as being proportional to the corresponding change in pressure* of this gas (hydrogen thermometer, § 242). Other temperature effects, for example the relative change of volume of a gas under constant pressure, or the relative expansion of mercury in glass, furnish thermometers which closely agree with the hydrogen thermometer, but a reduction to the latter is always necessary where great accuracy is required.

241. The Degree. Centigrade and Fahrenheit Scales. After the selection of a thermometric substance with its characteristic property there still remains the choice of the unit scale division and of a zero point. The unit for the measurement of temperature is called a *degree.*

In all scientific work the *centigrade scale* is used, in which the degree is defined as the one-hundredth part of the interval between the temperatures of melting pure ice and of the steam of water boiling under *normal atmospheric pressure* (76 cm. of mercury). The former temperature, often called simply the melting point (or freezing point) is marked 0°; the latter, the boiling point, 100°. These two points, which are of fundamental importance for the standardization of thermometers, are called the *fixed points* of the thermometer, and their difference in temperature the *fundamental interval.* All numerical data given in this book are expressed in degrees centigrade unless otherwise stated.

In the *Fahrenheit thermometers,* which are widely used in this country, the freezing point of water is at + 32 degrees, and the boiling point at + 212 degrees. Thus 180 parts of the Fahrenheit scale correspond to 100 of the centigrade scale. In order to reduce the reading on one of these two scales to the corresponding reading on the other we refer them to the same temperature, namely the melting point, and obtain the equation:

$$\frac{F - 32}{180} = \frac{C}{100}$$

or

$$F = \frac{9}{5}C + 32 \quad \text{and} \quad C = \frac{5}{9}(F - 32) \qquad (1)$$

It is evident that either scale may be used in a standard thermometer or in any other thermometer, based upon a different principle.

The centigrade scale is sometimes called the Celsius scale, though Celsius (1742) called the melting point 100 degrees and the boiling point zero. Fahrenheit (1714) himself did not employ the freezing and boiling points of water as fixed points, but the temperature of a mixture of water, ice and sal ammoniac, which he chose as zero point and blood heat which he called 96 degrees. Another thermometric scale derives its name from its inventor, Réaumur, who divided the fundamental interval into 80 degrees, choosing the freezing point of water as 0°.

242. The Constant Volume Gas Thermometer. In these thermometers the increase of temperature is defined as proportional to the relative increase in pressure of a gas whose volume is kept constant.

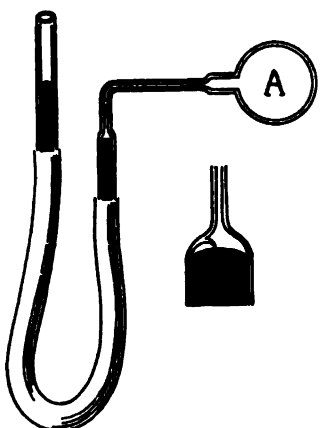


Fig. 133.

The *dry* gas is enclosed in a suitable vessel of glass or some other substance (glazed porcelain, platinum, or platinum-iridium) (see Fig. 133). A capillary tube connects the vessel with the manometer, which usually consists of a U-tube carrying near the end of the capillary a fine pointer. The gas is shut off from the outer air by mercury contained in this U-tube. In order to keep the volume of the enclosed gas constant, the mercury is always adjusted so that its surface on the side of the tube connected with the capillary just touches the pointer.

The pressure to which the gas is subjected at any temperature can easily be found from the barometric pressure and the difference in height of the mercury on the two sides of the manometer tube, a small correction being necessary on account of the change in volume of the containing vessel.

In the standard hydrogen thermometer the hydrogen gas is

under a pressure of 100 cms. of mercury at the temperature of melting ice but it will give accurate results, even if the initial pressure is smaller. Let P_0 be the pressure of the hydrogen when it is at the temperature of melting ice, and P_{100} its pressure when at the temperature of steam. The difference in temperature is 100°, and one degree is, therefore, according to § 240, defined as the rise in temperature that will produce an increase in pressure of $(P_{100} - P_0)/100$. Hence t degrees will produce an increase of pressure of $t(P_{100} - P_0)/100$; and, if the pressure at $t°$ above the lower fixed point be P_t,

$$P_t - P_0 = t\,\frac{P_{100} - P_0}{100} \tag{2}$$

or

$$t = 100\,\frac{P_{100} - P_0}{P_t - P_0}\ \text{degrees.} \tag{3}$$

From (2) we see that

$$\frac{P_t - P_0}{P_0 t} = \frac{P_{100} - P_0}{100 P_0} = a_p \tag{4}$$

Thus *the proportional increase of pressure per degree above zero* is a constant. This constant is called the coefficient of increase of pressure, or simply the pressure coefficient of hydrogen, and is denoted by a_p.

Hence from (4)

$$P_t = P_0(1 + a_p t). \tag{5}$$

The value for a_p (for hydrogen) is 0.0036624 per degree.

Since hydrogen diffuses at high temperatures through the walls of the containing vessel other gases, as nitrogen or air, which do not show this tendency, are frequently substituted for hydrogen. In ordinary laboratory work air is generally used. The pressure coefficients for these gases differ very slightly from that of hydrogen. For temperatures lying between 0 and 100 degrees the differences between the readings on the standard thermometer and those taken with the constant volume nitrogen or air thermometers amount only to a few hundredths of a degree, but at lower temperatures they become appreciable.

243. The Constant Pressure Gas Thermometer. The large ex-

pansion of air when heated was known to the ancient philosopher, Hero of Alexandria, and it is doubtless due to this striking property that the earliest methods for measuring temperature were based upon it. Galileo is probably the inventor of the "air-thermometer" which in its simplest form is represented in Fig. 134.

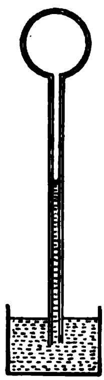


FIG. 134.

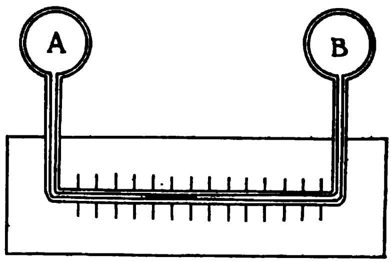


FIG. 135.

The gas in the bulb is shut off from the outside air by a colored liquid partly filling the narrow tube. The expansion is made visible by the displacement of the liquid column, while the pressure due to the atmosphere and the liquid remains nearly constant.

The construction of these thermometers has been considerably changed by a number of investigators but they are seldom used for precise measurements; they are, however, very sensitive indicators of temperature changes and valuable for demonstration purposes. A well known special form is the "differential thermometer" in which a slight difference in the temperatures of the bulbs is indicated by a relatively large displacement of the liquid enclosed in the narrow tube between them. (See Fig. 135.)

244. Liquid Thermometers. Since the gas thermometers are bulky and inconvenient for general use the expansion of liquids was early employed for temperature measurements. The liquid, usually mercury or alcohol, is contained in a glass reservoir connected with a capillary stem of uniform cylindrical bore.

245. Mercury-in-Glass Thermometers were first constructed by

Boullieau in 1659 and great improvements were made by Fahrenheit in 1724. Since that time these thermometers have been in general use. The principle of temperature measurement by means of these instruments is based upon the definition that a temperature change is proportional to the apparent increase in volume of the mercury. This apparent increase is the difference between the actual change in the volume of mercury and that of the containing vessel.

It is not to be expected that temperature differences defined in this way should agree exactly with those of the normal hydrogen thermometer, and indeed the errors are considerable for low and high temperatures. Moreover glass is not a well defined chemical substance, and thermometers made of different kinds of glass will not perfectly agree among themselves.

Such thermometers present also other practical disadvantages for measurements of precision. The volume of the bulb contracts slowly for a long period, which results in course of time in a gradual rise of the zero point. Moreover after being heated considerably, the bulb upon cooling will not immediately return to its original volume, but will at first be somewhat larger, this condition being followed by a slow contraction (depression of the zero point). On account of the hydrostatic pressure of the mercury upon the bulb different readings will be obtained with very sensitive instruments in a vertical and in a horizontal position. Often a part of the mercury thread extends outside of the body whose temperature is to be measured and a correction must be made for the difference in the temperatures of the bulb and of the exposed mercury column. (Stem correction.) These corrections, which must always be considered in accurate work, are, however, small and can frequently be entirely neglected, or need be known only approximately.

Since mercury freezes at —38.8° it cannot be used for the measurement of lower temperatures. It boils at 357° under atmospheric pressure but boiling can be prevented by an increase of pressure (§ 290). If the glass is very strong and the space above the mercury filled with a chemically inert gas,—nitrogen or carbon dioxide—the pressure of the gas will constantly increase as the mercury column rises and may be made sufficiently large to

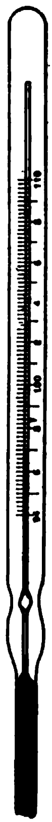


FIG. 136.

prevent boiling even at 500°. The range of mercury in glass thermometers extends thus from —39° to +500°, the upper limit being due to the fact that glass becomes soft at still higher temperatures. By using for the envelope fused quartz, which melts at about 2000°, mercury-in-quartz thermometers have been constructed which allow temperature measurements up to 700°.

246. Other Liquid Thermometers. For temperatures below 39° liquids must be used which have a low freezing point. Alcohol, which freezes at —111°, has been used generally for this purpose, but these thermometers show great irregularities. For this reason toluene is to be preferred, which also has the advantage of boiling above 100°, namely 111°, and thus permits a determination of both fixed points, and the measurement of temperatures as low as —80°. Pentane, and a liquid obtained by distilling petroleum ether and having a boiling point at about 20°, remain liquid even at —200° and can, therefore, be used for very low temperatures. Still other liquids have been proposed and used for special purposes. All such thermometers must, of course, be carefully calibrated by comparison with a standard.

247. Maximum and Minimum Thermometers. Maximum and minimum thermometers are instruments which register the highest and lowest readings reached during any time interval. Two types of maximum thermometers are widely used. They are both mercury-in-glass thermometers. In the first a thin iron rod with a small spring pressing against the wall of the tube is placed in the bore

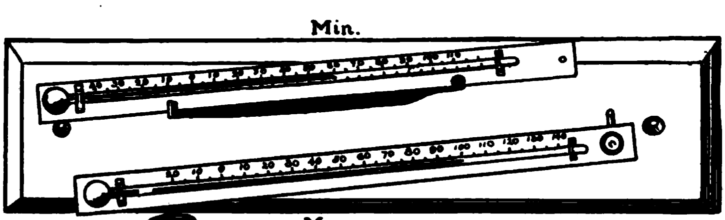


FIG. 137.

just *outside* the mercury thread. With rising temperature this rod is pushed forward by the mercury, but remains stationary when the thread recedes. The end of the rod nearest the bulb indicates the maximum temperature. The iron rod can be brought back to the mercury surface by means of a little magnet, or by shaking. In clinical thermometers the iron index is often replaced by a small section of mercury thread separated from the main thread by a little air.

In the second form there is a contraction in the bore of the glass near the bulb. With rising temperature the expanding mercury pushes through the constriction in the stem, but on contraction the thread breaks off at this point and the mercury remains in the stem giving directly the highest reading. The mercury thread is brought back to its proper position by shaking or centrifugal force. This principle is extensively used in clinical thermometers. (See Fig. 136.)

The minimum thermometer is usually an alcohol-in-glass thermometer, kept in a nearly horizontal position. A small glass rod is placed as an index in the bore *inside* of the alcohol. When the thread recedes on cooling its surface tension pulls the glass rod back with it, but the liquid flows past it when the temperature rises. The point of the index farthest from the bulb marks thus the minimum temperature. The rod is brought back to its original position by raising the bulb and allowing the index to slip forward to the surface of the alcohol.

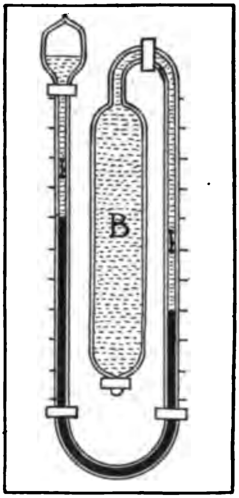


FIG. 138.

Fig. 137 shows the maximum and minimum thermometers used by the United States Weather Bureau.

In Six's form, Fig. 138, both thermometers are combined in one, both indices being small steel rods, pushed along the bore by the mercury in the U-tube and held by small springs; the rather large thermometer bulb, *B*, is filled with glycerine.

248. Other Methods of Temperature Measurement. Among the numerous methods for temperature measurement only those most frequently used will be mentioned here though a complete discussion of the underlying principles must be reserved for later chapters. All thermometers described below must be calibrated by comparison with the standard thermometer.

(*a*) Different solids in general expand differently with rise of temperature. If the flat sides of two metal strips of different material are soldered together, the compound strips will on becoming warmer curve towards the side of the metal least affected by the temperature rise or, if the strip is bent into a spiral spring, it will twist or untwist with a change of temperature. This principle is employed in Breguet's *strain thermometer* (see Fig. 139), and in the self-registering thermographs, which are used for the continuous registration of temperatures.

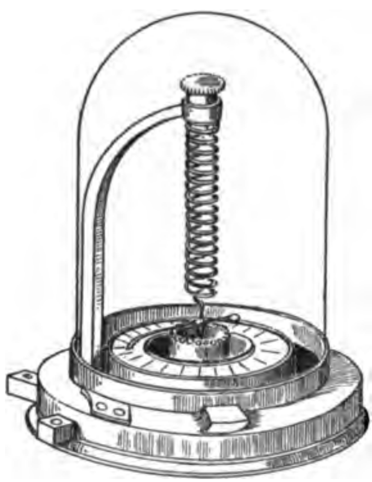


Fig. 139.

(*b*) The electrical resistance of a wire changes with the temperature (see "Electricity"). By measuring its resistance the temperature of the wire can be accurately determined. Siemens introduced the *resistance thermometer* in 1871, but it has been greatly improved by Callendar, who coils a fine wire of pure platinum around a mica frame and protects it by a suitable vessel. (Fig. 140.) These thermometers are well adapted for low temperature measurements.

(*c*) If the junctions of two dissimilar metals are kept at different temperatures a thermo-electromotive force is produced which can be measured by appropriate electrical means, for instance by the electrical current produced in a galvanometer circuit. Such a system of two wires is called a *thermo-element* or ther-

mocouple. For the determination of the temperature of a given medium one junction is placed in this medium while the other ends are usually placed in melting ice. The electromotive force is then measured and from it the difference in temperature between the two junctions calculated. LeChatelier's thermo-couples, or pyrometers, consist of one wire of pure platinum and one of platinum containing 10 per cent of rhodium. They allow measurements as high as 1700°.

FIG. 140.

(*d*) All bodies radiate energy into space. The intensity of the radiation increases very rapidly with increase of temperature and finally the bodies become luminous. It is well known that we can estimate roughly the temperature of a glowing solid from the nature of the light which it emits.

In recent years the laws governing ***radiation*** have been investigated and upon them important methods of measuring temperatures are based (*Optical Pyrometry*). These enable us to measure the temperature of bodies beyond our reach, by a study of their radiation; the student is referred for a more complete treatment to the section on radiation (see § 315).

249. The Zero of the Gas Scale. From equation (5) in § 242 it follows that if P_t is placed equal to zero

$$t = -\frac{1}{\alpha_p} = -\frac{1}{0.0036624} = -273° \qquad (6)$$

This temperature is called the "zero of the gas scale," or "absolute zero,"* and temperatures measured from this point "absolute temperatures." In what follows these will be denoted by T. To express a given reading on the Centigrade scale in "degrees absolute" 273 degrees have to be added, or

$$T° = t° + 273° \qquad (7)$$

Equation (5) then becomes,

$$P_t = \alpha_p P_0 T, \qquad (8)$$

or, since α_p and P_0 are constant, the pressure is proportional to the absolute temperature. Absolute zero is, therefore, the tem-

* For a truly absolute scale, *i. e.*, one which is independent of any particular substance, see § 329.

perature at which hydrogen gas would exert no pressure supposing that it remains a gas even at the lowest temperature, a supposition which is not in agreement with the facts, since hydrogen has been liquefied and even frozen at very low temperatures (§ 305). According to the molceular theory no pressure in a gas would mean no motion of the molecules. A gas would, therefore, contain no heat energy at absolute zero.

EXPANSION.

250. Linear Expansion of Solids. Let a metal bar be clamped at one end and the other end rest upon a sewing needle placed at right angles to the bar and free to roll upon a smooth surface (glass). The needle carries a long, light pointer, which is counterbalanced so that free rotation may not be hindered. When the bar is heated, for instance by a Bunsen burner, it slightly expands and the advancing free end will rotate the needle. This motion is magnified by the pointer. Upon cooling, contraction of the bar, and a return of the pointer to its original position will result. Since the expansion of solids with change of temperature is quite small, very refined methods must be

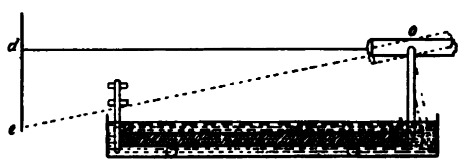


FIG. 141.

used for its measurement. Fig. 141 represents a schematic outline of the apparatus employed by Laplace and Lavoisier for this purpose. One end *a* of the bar whose expansion is to be measured is held in a fixed position, while the other end *b* presses against a vertical lever which turns around a horizontal axis *o*. A telescope, rigidly connected with the lever and therefore turning through the same angle, is focused on a distant scale at *d*. The bar is kept in an oil or water bath. Let the change of length of the bar upon heating be *bc* and the change of the reading of the scale *de*. Since *bco* and *deo* are similar triangles, the expansion of the bar is easily calculated, if the length of the lever and the distance of the axis of rotation from the scale are known.

In general the linear distance between any two points of a solid

increases with the temperature and the increase, $L_2 - L_1$, is nearly proportional to the differences in temperature $t_2 - t_1$, or

$$L_2 - L_1 = a_l L_1 (t_2 - t_1), \tag{9}$$

and

$$a_l = \frac{L_2 - L_1}{L_1 (t_2 - t_1)} \tag{10}$$

where a_l is called the coefficient of linear expansion in the direction L and may be defined as *the relative change in length per degree.* Usually 0° is chosen for the lower temperature and equation (9) may then be written,

$$L_t = L_0 (1 + a_l t) \tag{11}$$

where L_0 is the length at 0°, and t the higher temperature.

For a temperature below 0° t must of course be taken with the negative sign and if a_l is positive L_t is smaller than L_0.

Strictly speaking the length is not a linear function of the temperature and when plotted will not be a straight line. a_l itself depends, therefore, upon the temperature, so that

$$a_l = a' + a''t + a'''t^2 + \cdots \tag{12}$$

and equation (10) becomes,

$$L_t = L_0 (1 + a't + a''t^2 + \cdots) \tag{13}$$

Generally the first two coefficients are sufficient for an accurate expression of the curve. a_l as defined by equation (9) is called the *mean* coefficient between the temperatures t_2 and t_1. Frequently the mean coefficient between 0° and 100° is given which is approximately equal to $a' + 100a''$.

For copper a' is 0.0000167 and $a'' = 0.000000004$; the mean coefficient between 0° and 100° is, therefore, 0.0000171 per degree.

In the case of isotropic bodies or cubical crystals the dilatation is the same in all directions; in uniaxial crystals (rhombic and hexagonal) there is one axis of symmetry with a definite coefficient of linear expansion in this direction and a different one at right angles to it. In still other crystals, there are three principal axes of dilatation at right angles to each

other, each with its characteristic coefficient. The expansion in any direction depends upon the orientation of the length to be measured with respect to these axes. From this it follows that in general, with the exception of cubical crystals, the angles made by the plane faces of crystals will change with the temperature.

LINEAR EXPANSION OF SOLIDS.

Substance.	a'	a''	Substance.	a'	a''
Copper	16.7×10^{-6}	0.4×10^{-8}	Jena Glass 59^{III}	5.6×10^{-6}	0.2×10^{-8}
Gold	13.6	1.1	Jena Glass 16^{III}	7.7	0.3
Iron	11.7	0.5	Porcelain (Berlin)	2.7	0.3
Lead	27.3	0.7	Quartz ∥ axis	7.1	0.8
Nickel	13.5	0.3	Quartz ⊥ axis	13.2	1.2
Platinum	8.9	0.5	Quartz fused	0.39	0.17
Silver	18.3	0.5	Ice	52	
Zinc	27.4	2.3	Hard Rubber	80	
Brass	17.9	0.5	Oak ∥ grain	4.9	
Nickel steel (36% Ni)	0.9	0.1	Oak ⊥ grain	54.4	

251. Applications. The expansion of solids is of great practical importance. In the laying of railway rails or in the construction of metal bridges, for instance, allowance must be made for expansion and contraction due to temperature changes. If a piece of metal is to be shrunk upon a frame or another piece of metal, it is first heated sufficiently so that it will loosely fit the frame. Upon cooling it becomes firmly attached, the pressure exerted upon contraction being very large.

If glass is suddenly heated or cooled it breaks to pieces, since the temperature differences between the outer and inner portions produce larger strains than the glass can support.

Quartz crystals when strongly heated burst into innumerable small pieces owing to the unequal expansion in different directions, but after they have been fused to a homogeneous mass the coefficient of expansion is very small (see table), and on this account vessels made of fused quartz (quartz glass) may be heated to red heat and at once plunged into cold water without being broken.

In many measurements it is of the utmost importance that the dimensions of an instrument should be independent of temperature changes.

A well-known method for compensating the expansion of a pendulum is that used on the gridiron pendulum, figure 142. The dark lines represent iron rods and the light lines, zinc rods. The expansion of the iron lengthens the pendulum, but that of the zinc rods which have a larger coefficient of linear expansion will raise the pendulum bob sufficiently to keep the effective length of the pendulum constant.

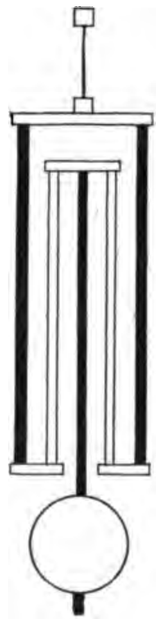


FIG. 142.

The balance wheels of chronometers are compensated for temperature effects by making the rim of the wheel of two metals, the one having the larger coefficient of expansion being on the outside. A rise in temperature diminishes the elasticity of the spring and slightly increases the diameter of the wheel. Both of these effects would increase the period, but the unequal expansion of the two metals bends the rim into a smaller circle, and thus decreases the moment of inertia of the balance wheel and the period remains constant. A similar application for the construction of the strain thermometers was mentioned in § 248.

A very valuable metal is an alloy of nickel and steel. A steel containing about 36 per cent. of nickel has a coefficient of linear expansion of only 0.0000009 per degree or one tenth that of glass. Invar is a nickel steel in which, by mechanical treatment, the coefficient has been still further reduced. It is used extensively for the construction of standards of length, steel tapes, pendulums, etc.

252. Cubical Expansion of Solids. In a manner similar to that of the last article the relation between the volume of a body at the temperatures $t°$ and 0° may be expressed by the equation

$$V_t = V_0(1 + a_v t) \qquad (14)$$

a_v is called the coefficient of cubical expansion and can easily be found from the coefficient (or coefficients) of linear expansion.

Let the lengths of the edges of an isotropic rectangular parallelo-

piped at 0° be L_1, L_2, L_3; then at this temperature

$$V_0 = L_1 L_2 L_3 \tag{15}$$

At $t°$ the lengths have increased to $L_1(1 + a_l t)$, etc., and the volume to

$$\begin{aligned} V_t &= [L_1 \cdot (1 + a_l t)] \cdot [L_2 \cdot (1 + a_l t)] \cdot [L_3 \cdot (1 + a_l t)], \\ &= L_1 L_2 L_3 (1 + 3a_l t + 3a_l^2 t^2 + a_l^3 t^3) \end{aligned} \tag{16}$$

but since a_l is small the last two terms in the parentheses may be neglected in comparison with $3a_l t$, and we may write

$$V_t = V_0(1 + 3a_l t) \quad \text{or} \quad a_v = 3a_l \tag{17}$$

In the case of anisotropic bodies a_v equals the sum of the coefficients of linear expansion, a_1, a_2 and a_3 parallel to the principal axes of dilatation, or

$$a_v = a_1 + a_2 + a_3 \tag{18}$$

which simplifies in the case of uniaxial crystals to

$$a_v = a_1 + 2a_2 \tag{19}$$

As a rule solids expand when heated, *i. e.*, the coefficients of expansion are positive. Notable exceptions are iodide of silver which contracts between $-60°$ and $+142°$, cuprous oxide which has its maximum density at 4.1°, diamond and fused quartz with maximum densities at $-39°$ and $-60°$ respectively.

A caoutchouc (rubber) tube stretched to twice its original length, or more, shortens considerably when steam is passed through it. Though the coefficient of linear expansion parallel to its length is negative, its volume increases, showing that the coefficient at right angles to the tube is positive and that caoutchouc becomes upon stretching an anisotropic body.

253. Expansion of Liquids. In liquids (and gases) we have to consider only volume expansion and, with the exception of water and aqueous solutions, the coefficient is always positive and larger than that of solids, but it changes considerably with temperature, so that for accurate determinations the third or even the fourth term of the empirical formula

$$V_t = V_0(1 + \alpha t + \beta t^2 + \gamma t^3 + \cdots) \tag{20}$$

must be taken into account.

As in § 250 the mean coefficient between the temperatures t_1 and t_2 is given by

$$\alpha_v' = \frac{V_2 - V_1}{t_2 - t_1} \frac{1}{V_1} \tag{21}$$

We must distinguish between the real expansion and the apparent expansion of a liquid contained in a vessel. When a hollow body is heated the hollow space expands just as much as if it were filled with the same substance as that of the walls. Let the liquid occupy at 0° the volume V_0 of the vessel, whose graduation is supposed to be correct at this temperature. After vessel and liquid are heated to t°, the liquid rises in the vessel to a mark indicating a volume V_a, and this apparent volume is, therefore,

$$V_a = V_0(1 + \alpha_a t), \tag{22}$$

where α_a is the apparent coefficient of expansion of the liquid.

However, since the vessel has also expanded, the real volume of V_a at t° is, if we call α_g the coefficient of expansion of the vessel,

$$V_t = V_a(1 + \alpha_g t) \tag{23}$$

and for the liquid:

$$V_t = V_0(1 + \alpha_v t), \tag{24}$$

where α_v is the real coefficient of expansion of the liquid. By eliminating the volumes from the last three equations, we obtain

$$1 + \alpha_v t = (1 + \alpha_a t)(1 + \alpha_g t)$$

or neglecting the small quantity $\alpha_a \cdot \alpha_g t$ in comparison with $\alpha_a + \alpha_g$

$$\alpha_v = \alpha_a + \alpha_g \tag{25}$$

The real coefficient of expansion can therefore be determined, if that of the vessel is known, by observing the change of the apparent volume with the temperature.

A method independent of the coefficient of expansion of the vessel is based upon the determination of the densities of the

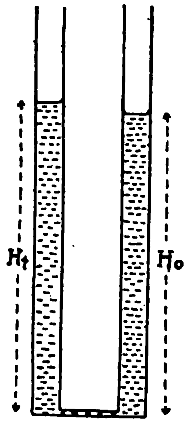


FIG. 143.

liquid. If V_0, V_t and d_0, d_t are the volumes and densities at 0° and t° respectively, we have for a given mass, M,

$$M = V_0 d_0 = V_t d_t \qquad (26)$$

But

$$V_t = V_0(1 + a_v t) \qquad (27)$$

and

$$\frac{d_t}{d_0} = \frac{V_0}{V_t} = \frac{1}{1 + a_v t} \qquad (28)$$

Let the liquid be in two communicating tubes (Fig. 143), which are kept at 0° and t° respectively, and let the heights above the common plane be H_0 and H_t. Since the pressure upon this plane of area A is the same on both sides,

$$AH_0 d_0 g = AH_t d_t g \qquad (29)$$

and

$$\left.\begin{aligned} \frac{H_t}{H_0} &= \frac{d_0}{d_t} = 1 + a_v t. \\ H_t &= H_0(1 + a_v t) \end{aligned}\right\} \qquad (30)$$

from which a_v of the liquid can be calculated.

Regnault's (1810–1878) classical determinations of the dilatation of mercury were made with an apparatus based upon this principle.

CUBICAL EXPANSION OF LIQUIDS UNDER ATMOSPHERIC PRESSURE.

Substance.	α	β
Mercury	181.8×10^{-6}	0.78×10^{-8}
Alcohol	1020	200
Ether	1480	350
Benzol	1160	223
Pentane	1465	309

The coefficient of expansion of liquids increases considerably with the temperature and becomes in some cases larger than that of gases, especially in the case of liquefied gases. Liquid carbon dioxide, for example, in-

creases by half its volume when heated in a closed tube from 0° to 30°, the mean coefficient between these temperatures being 0.017.

The expansion of liquids is not independent of the pressure; it decreases, except in the case of water, with an increase of pressure. For ether under atmospheric pressure the mean coefficient between 0° and 20° is 0.00156, for 100 atmospheres 0.00144, for 1000 atmospheres 0.0009.

254. Expansion of Water. Water forms a remarkable exception to the general rule that liquids expand with rising temperature. Its maximum density is at 4°, and below this temperature

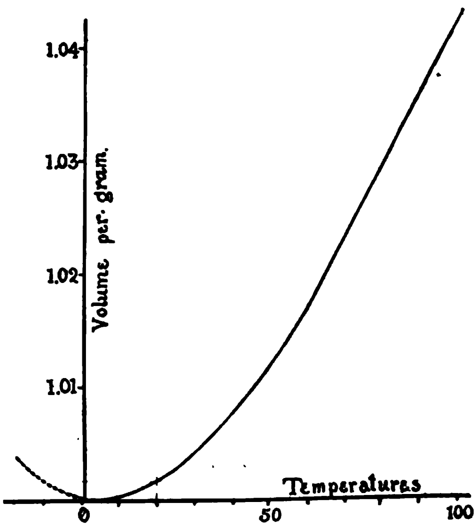


FIG. 144.

it expands again. In Fig. 144 the volume of one gram of water is plotted as a function of the temperature, and in the table some numerical values are given.

EXPANSION OF WATER UNDER ATMOSPHERIC PRESSURE.

Temp.	Vol.	Temp.	Vol.	Temp.	Vol.
—10°	1.00186 c.c.	4°	1.00000 c.c.	20°	1.00177 c.c.
— 5	070	5	001	30	435
0	013	6	003	40	782
1	007	7	007	50	1.01705
2	003	8	012	80	2899
3	001	10	027	100	4343

At 0° water changes to ice, a process accompanied by considerable expansion (§ 277). The curve for ice is practically a straight line, the ice contracting regularly with decreasing temperature. By avoiding mechanical disturbance water can be cooled considerably below the freezing point; the expansion of this *undercooled* water is represented in the figure by the dotted line.

If the surface of a lake is slowly cooled the surface layers become heavier and slowly settle to the bottom being replaced by warmer portions. This continues until the whole mass has reached a temperature of 4°. On further cooling the surface layer becomes lighter than the water below and remains at the top, until the freezing point is finally reached; for this reason water freezes from the surface downward instead of from the bottom up.

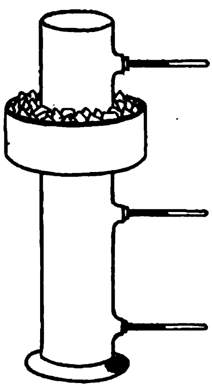


Fig. 145.

Hope's apparatus, Fig. 145, is designed to illustrate this process. The temperatures of the different layers of the water in the cylindrical vessel is measured by thermometers inserted at different heights at the side. Cooling is produced by a freezing mixture of ice and salt surrounding the upper part of the vessel.

The exceptional behavior of water may also be studied by observing the relative expansion of water in a glass vessel consisting of a large bulb with a small, graduated capillary tube (dilatometer). Upon cooling the meniscus will fall, but, after reaching a lowest point, it will rise again. On account of the expansion of the vessel (see equation 25), the lowest point is reached when the water is at about 5°, for $a_a = 0$ if $a_g = a_v$.

The above applies only if the pressure does not differ much from atmospheric pressure. If it is considerably increased the maximum density will be found at lower temperatures, for a pressure of 41.6 atmospheres it is at 3.3° and for 145 atmospheres at 0.6°.

The presence of substances dissolved in water also lowers the temperature of maximum density in a marked degree. The following table shows the influence of common salt, NaCl.

INFLUENCE OF SALT UPON MAXIMUM DENSITY OF SOLUTION.

Salt Dissolved in Per Cent.	Temperature of Maximum Density.
1	+ 1.77
2	— 0.56
4	— 5.63
8	—16.62

The maximum density for sea water would be at — 3.67°, but before this point is reached freezing will occur, under the existing conditions, currents be set up, and the equilibrium disturbed. At great ocean depths the temperature of the water is about 2.5° and remarkably uniform for large areas.

255. Expansion of Gases. Law of Gay Lussac. As has already been pointed out (§ 243) the expansion of air served as the first means for temperature measurements. While measuring the expansion of different gases Gay Lussac found in 1802 that under constant pressure "different gases possess the same coefficient of expansion." This is known as Gay Lussac's law or Charles' law. Charles, according to Gay Lussac's statement, discovered this property of gases in 1787, but never published an account of his experiments.

More accurate measurements have, however, shown that this law is by no means exact, and holds approximately only for gases which closely obey Boyle's law, while, especially in the case of gases which can easily be liquefied—for example in the last three gases in the table below—the coefficient is considerably larger.

COEFFICIENTS OF EXPANSION AND PRESSURE FOR GASES.

Substance.	a_v (Constant Pressure) = 76 cm. Hg.	a_p (Constant Volume) P_0 = 76 cm. Hg.
Hydrogen	0.003660	0.003663
Helium	3663	3664
Nitrogen	3671	3668
Air	3671	3665
Carbon dioxide	3710	3690
Cyanogen	3877	3682
Sulphur dioxide	3903	3670

256. The Gas Law. Assuming hydrogen to be an ideal gas, *i. e.*, one which obeys Boyle's law, the coefficient of cubical expansion

in the equation

$$V_t = V_0(1 + \alpha_v t) \tag{31}$$

must be equal to the pressure coefficient α_p (§ 242). For from the definition of temperature we have for a given mass of hydrogen of constant volume, when heated to the temperature $t°$,

$$P_t = P_0(1 + \alpha_p t). \tag{32}$$

Now let a mass of gas which occupies at 0° a volume V_0 under a pressure P_0 be heated to $t°$ while its volume is kept constant. Then

$$P_t V_0 = P_0 V_0(1 + \alpha_p t) = C_1 \tag{33}$$

On the other hand heating the same mass to the same temperature under constant pressure, P_0, we have

$$P_0 V_t = P_0 V_0(1 + \alpha_v t) = C_2 \tag{34}$$

from which it follows that, according to Boyle's law, since $P_t V_0$ and $P_0 V_t$ are the product of pressure and volume for the same temperature, t, and therefore equal,

$$\alpha_v = \alpha_p \tag{35}$$

Since $t = T - 1/\alpha_p$ (§ 249) it follows in general since the product $PV(=P_t V_0 = P_0 V_t)$ of a perfect gas is a constant independent of pressure or volume, that

$$PV = \alpha_p P_0 V_0 T = R_1 T \tag{36}$$

where T is the absolute temperature and R_1 a constant for a given mass of gas. Denoting by P' and V' the pressure and volume at any other temperature T', we obtain the important relation

$$\frac{PV}{T} = \frac{P'V'}{T'} \tag{37}$$

Equation (36) also shows that at absolute zero there is no energy (see § 227) in a gas. This equation is frequently called the gas law but it can easily be made more general.

(1) The volume V of a gas is proportional to its mass, or $V = Mv$ where v is the specific volume, or the volume of unit mass of the gas. Then

$$Pv = R_2T, \tag{38}$$

where $R_2 = R_1/M$ is a constant for a given gas.

(2) According to Avogadro's hypothesis equal volumes of different gases contain under the same pressure and at the same temperature an equal number of molecules, or their masses are proportional to the molecular weights of the gases. The volume V' occupied by one *gram-molecule, i. e.*, a number of grams equal to the molecular weight of the substance, for example 2 grams of hydrogen or 32 grams of oxygen, is therefore the same for all gases under a given pressure and at the same temperature, or $V' = nv$, where n is the number of grams equal to the molecular weight.

Then from (38)

$$PV' = nR_2T = RT. \tag{39}$$

The constant $R = nR_2 = n\dfrac{R_1}{M}$ is *the same for all gases.*

For example, the density of oxygen ($n = 32$ gr.) at 0° ($T = 273°$) and under atmospheric pressure is 0.0014291 gr./cm.^{3}; $v = 1/d = 699.74$ cm.^{3}/gr. and $V' = 22{,}390$ cm.^{3}

Since $P = 1012630$ dynes/cm.^{2}

$$R = \frac{PV'}{T} = \frac{22390 \cdot 1012630}{273} = 8.305 \times 10^7 \frac{\text{ergs}}{\text{degree}}$$

or in general

$$PV' = 8.305 \times 10^7 T \text{ ergs.} \tag{40}$$

For gases which do not follow Boyle's law the above conclusions are only approximately correct. With the exception of hydrogen and helium PV decreases with an increase of pressure under ordinary conditions (§ 223); therefore C_1 in equation (33) must be smaller than C_2 in (34) or $a_p < a_v$. In the case of hydrogen PV decreases with increasing pressure and $a_p > a_v$.

Since all gases approach the ideal state with a decrease of pressure, the coefficients will approach equality as the pressure is decreased. In the case of hydrogen the effect of pressure is, however, very small.

Regnault derived the same conclusions from his experiments on the expansion of gases, and found also that with the exception of hydrogen the coefficient of expansion as well as of pressure increase with the pressure.

CONVECTION OF HEAT.

257. Fluids. When a portion of a liquid or gas is heated it becomes less dense than the surrounding mass and will rise, being replaced by denser, cooler portions. Thus currents are set up in the mass of the fluid. The heating of water in a flask with the

heat applied to the bottom of the flask furnishes a good example of this. The water in the center rises and a downward current is found at the sides of the vessel.

The heated particles are constantly carried from points of higher temperature to points of lower temperature and there they will give up part of their heat energy to bodies with which they come into contact.

We speak therefore of convection of heat or transference of heat by convection, which is characterized by a translatory motion of heated matter.

258. Liquids. Convection of heat by water is used for the heating of buildings. The principle of the method is represented in Fig. 146. The water heated in the boiler rises through the pipe *BC*, circulates through the building, gives up heat in the radiators *R* and returns finally at a lower temperature to the boiler. At the highest point of the system is an open tank *D* to allow for the expansion of water during the heating.

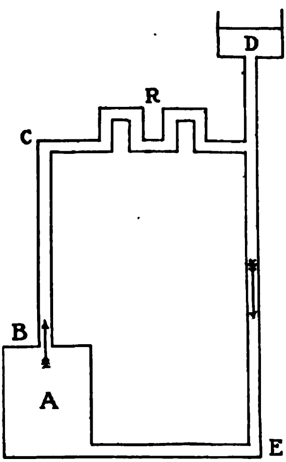


Fig. 146.

Some ocean currents are probably convection currents. The Gulf Stream, for example, originates near the equator. The heating of the ocean produces a rise in level and the water will flow off towards the lower levels farther north, being replaced by an under current of cold water.

259. Gases. Convection of heat by means of air is used for warming houses by hot air furnaces. The heated air is led directly through the registers into the rooms, being replaced in the furnaces by cold air from the outside.

Convection currents also serve for ventilation. The draft in chimneys is due to them, the air necessary for the combustion of coal usually being supplied by them at the same time. Where, however, a stronger draft is needed, fans or blowers are used in addition.

A well known phenomenon due to convection currents in the atmosphere is easily observed on hot summer days. Heated air is constantly rising from the surface of the earth and is replaced by downward currents of cooler air from a short distance above. If this occurs the outlines of distant objects appear to be continually undulating because the power of refraction for light differs for warm and cold air. This can easily be shown by holding a gas flame in the path of a beam of light from a projection lantern. The quivering outlines of the heated upward stream of air can be seen distinctly on the screen. The twinkling of the stars is also due to convection currents in the air at greater heights.

The most important examples of convection currents are the winds. Near the sea coast, especially in tropical regions, the wind is a sea breeze during the day, since the air moves from the cooler ocean to the land above which the heated air rises; during the night it is a land breeze because the water is then warmer than the land and the direction of the movement of the air is reversed.

The trade winds are similar phenomena on a larger scale, and are due to equatorial heating, which produces an upward motion near the equator and a constant influx of surface air from north and south. Since the air currents move towards regions which have a larger velocity in the west-east direction, they appear to come from a northeasterly direction in the northern hemisphere and from a southeasterly direction in the southern hemisphere.

The upper air currents flowing off towards the poles strike the surface of the earth again at a latitude of about 35° and the one on the northern hemisphere appears as a southwest wind, since it comes now from regions having a larger velocity towards the east than that of the surface of the earth where it strikes. The mild climate of Western Europe is largely due to the prevalence of winds from the West.

Convection currents in the atmosphere are doubtless aided by the presence of water vapor which is lighter than air. A mixture of air and water vapor under a given pressure is therefore less dense than air alone (see § 284) and the presence of water vapor increases the difference in density between the warm, vapor-laden and the cooler, drier portions of the atmosphere.

To avoid loss of heat by convection, incandescent lamp globes

are exhausted. The difference thus produced may be shown by the following experiment (Andrews' experiment). A thin platinum wire is stretched along the axis of a glass cylinder, which allows exhaustion of the air. Send an electric current through the wire and heat it to a dull red. Upon exhaustion the wire becomes much hotter and brighter, because now it does not lose a large part of its heat by convection. But if hydrogen, whose molecules move faster than those of air, is introduced in the cylinder the cooling due to convection becomes so large that the wire carrying the same current as before remains dark.

CALORIMETRY.

260. The Calorie. When a vessel containing cold water is placed over a flame, the temperature of the water rises, and we say that heat passes from the flame to the water; but if twice the amount of water is taken, the temperature rises only half as fast, though there is every reason to believe that the same amount of heat as before passes into the water during the same time. On the other hand, if equal masses of mercury (or any other liquid) and water be heated by the same flame, the temperature of the mercury rises considerably faster than that of the water.

Apparently heat as a physical quantity may be measured by its effect upon the temperature of the body which it enters, but, in the selection of its unit, the substance used for such measurement, its mass, and its change of temperature must be carefully specified.

Early in the history of heat measurements, during the time when the caloric theory flourished, it was agreed to choose as the *unit of heat the quantity of heat which changes the temperature of one gram of water one degree.*

This unit is called the "calorie." It is entirely independent of any interpretation as to the nature of heat and it is still used for all experimental work in which purely thermal processes are considered.

261. Heat Capacity of a Body and of a Substance. Experiments have shown that a quantity of heat, when added to a given body, is proportional to the rise in temperature of this body, or

$$H = C(t_2 - t_1) \tag{41}$$

C, i. e., the ratio between the quantity of heat and the temperature interval, is called the heat capacity *of the body* and is numerically equal to the quantity of heat needed to change the temperature *of the body* one degree.

Further, the heat capacity is proportional to the mass of the body. Thus equation (41) may be written

$$H = cM(t_2 - t_1) \tag{42}$$

and

$$c = \frac{C}{M} \tag{43}$$

c is characteristic for any given substance and is therefore called the heat capacity *of the substance.* It is numerically equal to the quantity of heat required to raise the temperature of unit mass of the substance one degree. It differs for different substances and for modifications of the same substance, for example, diamond, charcoal and graphite. It also depends to a slight extent upon the temperature (§ 265) and upon the mechanical treatment which the body has received; thus soft copper has a different heat capacity from hard rolled copper. From a comparison of the definition of the calorie with equation (42) it is evident that the heat capacity of water is unity.

Strictly speaking, the heat capacity of water is not a constant for different temperatures. Rowland was the first to show that it decreases from 0° to about 30° and increases again from that point on. His results have been corroborated by other experimenters, though some of them found a slightly different temperature for the minimum, namely 40°. The variation at this point is, however, extremely small and difficult to determine.

For accurate work it is necessary to state the temperature interval for the exact definition of the calorie. No general agreement has been reached by physicists. It seems best to choose the interval between 15° and 16°, especially since in this case the calorie would be equal to the one-hundredth part of the heat required to raise the temperature of one gram of water from 0° to 100°, *i. e.,* the mean heat capacity of water between the fundamental points.

Frequently the "large calorie," *K,* is used, especially for thermochemical measurement (§ 267). This unit equals 1000 calories, 1 kg. of water being chosen instead of 1 g.

15

In engineering practice the "British thermal unit, B. T. U.," is employed. It is the heat required to raise the temperature of one pound of water one degree Fahrenheit.

262. Specific Heat. After the value of the heat capacity of water has been chosen as unity, the quantity of heat necessary to change the temperature of a given mass of any substance can easily be calculated if the ratio between the heat capacity of the substance and that of water is known. This ratio is called the specific heat and is numerically equal to the heat capacity of the substance. For this reason usually no distinction is made between these two quantities, and specific heat is often used to mean the heat capacity of a substance.

263. The Method of Mixtures. The best known method for determining the specific heat of a substance is the method of mixtures. This is based upon the experimental fact that, when two or more bodies, originally at different temperatures, are placed in thermal contact, and the heat exchanges take place exclusively between these bodies, the quantities of heat lost by one part of the system equals the quantity of heat gained by the other. (Principle of equal heat exchanges.)

For example, let a substance of mass M_1, and heat capacity c_1, be heated to the temperature t_1 and then plunged into M_2 grams of water of a temperature $t_2 < t_1$. Let the final temperature of the mixture be t.

The above principle gives then

$$c_1 M_1 (t_1 - t) = c_w M_2 (t - t_2) \tag{44}$$

and the specific heat, s, is

$$s = \frac{c_1}{c_w} = \frac{M_2}{M_1} \frac{(t - t_2)}{(t_1 - t)}. \tag{45}$$

The water is contained in a calorimeter, *i. e.*, a vessel with a stirrer and a thermometer. The temperature of these substances also rises from $t_2°$ to $t°$ so that in fact a larger amount of heat is absorbed, than would be by the water alone. This additional amount is

$$\Sigma c M (t - t_2) = M' (t - t_2) c_w \tag{46}$$

M' is called the *water equivalent of the calorimeter,* because it is

equal to a mass of water which, as far as heat exchanges are concerned, has the same effect as the calorimeter. It is easily calculated if the masses and specific heats of the vessel, stirrer and thermometer are known.

Equation (45) becomes then

$$s = \frac{M_2 + M'}{M_1} \frac{(t - t_2)}{(t_1 - t)}. \tag{47}$$

If the body whose specific heat is to be determined reacts chemically with water, a chemically inert liquid of known specific heat $s' = (c'/c_w)$ may be employed and the relative specific heat of the body with respect to this liquid, $s'' = (c_1/c')$ be found; then

$$s = \frac{c_1}{c_w} = \frac{c_1}{c'} \cdot \frac{c'}{c_w} = s'' \cdot s' \tag{48}$$

Frequently on account of heat exchanges with the surroundings heat is lost or gained by the system in the calorimeter. A correction for this, called correction for radiation, must be applied for accurate measurements. For this the students are referred to laboratory manuals.

Specific Heats of Solids and Liquids.

Substance.	s	Substance.	s	Substance.	s
Carbon (Diamond)	0.153	Mercury	0.033	Quartz	0.188
Carbon (Graphite)	0.199	Nickel	0.109	Jena Glass 16^{III}	0.199
Copper	0.094	Platinum	0.032	Basalt	0.21
Gold	0.031	Silver	0.056	Ice	0.505
Iron	0.116	Tin	0.055	Alcohol (16° to 30°)	0.602
Lead	0.031	Zinc	0.094	Ether (at 30°)	0.547

264. Specific Heat of Gases. Let a gas be enclosed in a cylinder, one end of which is formed by a movable piston on which a pressure P acts; when the gas expands it does work which equals $P\Delta V$, where ΔV is the increase of volume of the gas. Whenever a gas expands against pressure, it does external work. If a gas is expanded without heat being supplied its temperature will be lowered, since energy, equal to the external work done, must be supplied by the gas itself and is taken from the store of heat energy it contains. Thus if a gas is heated without change of volume, a smaller amount of energy is needed for a given rise in temperature than when it is allowed to expand and perform external work at the same time.

The heat capacity and specific heat of a gas are therefore indeterminate quantities unless the exact conditions under which the heating takes place, are specified. Similar reasoning would apply to solids and liquids, but their expansion is generally so small that its effect can be neglected.

We distinguish specific heat at constant volume, s_v, and the more important specific heat under constant pressure, s_p. The latter is much more easily measured and may be found, using the method of mixtures, by passing a measured quantity of the heated gas through a long coiled tube immersed in a calorimeter.

The two specific heats of gases stand in a definite relation to each other. Their ratio is of importance for the calculation of the velocity of sound in gases (see Sound). A further discussion of these quantities must be postponed to a later section (see thermodynamics, § 320).

SPECIFIC HEATS OF GASES.

Substance.	S_p	S_v	$\gamma = \frac{S_p}{S_v}$
Air..	0.237	0.167	1.41
Hydrogen	3.410	2.418	1.41
Oxygen	0.217	0.147	1.41
Water vapor (100°)	0.421	—	1.30
Carbon dioxide	0.203	—	1.30
Alcohol (108° to 220°)	0.453	—	1.13
Ether (25° to 111°)	0.428	—	1.08

265. Variation of Specific Heat with Temperature and State. The specific heat increases in general with the temperature (for water see § 261) and can be represented by the empirical formula

$$s = a + bt + ct^2 + \cdots \qquad (49)$$

where $a, b, c, \cdots$, are constants.

The change with temperature is smaller for solids than for liquids but becomes more marked in the neighborhood of the melting point. For liquids the specific heat is larger than for either the solid or gaseous state.

Regnault found that the specific heat under constant pressure of a gas approaching the ideal state, *i. e.*, one for which the gas law

holds (§ 256) depends neither on temperature nor pressure, and, if referred to equal volume and the same pressure, is the same for all such gases. Recent investigations have shown that this rule holds only for a limited number of diatomic gases, as hydrogen, oxygen, nitrogen and carbon monoxide. In general it is not true. The specific heat of easily condensible gases increases with the temperature and pressure.

266. Law of Dulong and Petit. It is quite arbitrary to refer the numerical values of heat capacities to unit mass; a law of a general nature can only be found by choosing a more natural unit, for example, the mass of a gram-atom or gram-molecule.

In 1819 Dulong and Petit discovered the following law, "The atomic heat, *i. e.*, the product of the specific heat into the atomic weight, is the same for all elementary solid substances." The number thus obtained is about 6.

Since the specific heat varies considerably with the temperature this rule can hardly be expected to be rigorously true. Carbon, boron and silicon, in which the temperature effect is especially pronounced, are notable exceptions to the general rule, but agree better with it at high temperatures.

It cannot, however, be denied that we have in this law more than an accidental coincidence and that some general law of nature is hidden behind it. Assuming that masses of different substances, proportional to the atomic weights, contain an equal number of atoms this law suggests that all atoms have the same heat capacity.

Other investigators have attempted to extend Dulong and Petit's law to chemical compounds and have discovered laws of more or less general application. For these the student is referred to textbooks of physical chemistry.

267. Thermochemical Measurements. An important problem in physical chemistry is the measurement of the quantity of heat developed or absorbed during chemical reactions.

These reactions are usually represented by equations, in which the chemical symbols not only serve to characterize the substance but stand also for a definite mass, namely, that corresponding to the atomic or molecular weight. Thus

$$\mathrm{Na} + \mathrm{Cl} = \mathrm{NaCl} \qquad (50)$$

means that one gram-molecule, *i. e.*, 58.5 g. of sodium chloride, is formed by the combination of 23.05 g. of sodium and 35.45 g. of chlorine.

But from a physical point of view the equation is incomplete, since during the chemical reaction a large amount of heat, namely, 20,400 calories, is developed. For this reason the meaning of the chemical symbols must be enlarged so as to include the energy contained in the gram-atom or gram-molecule of a substance, and we would write equation (50) now in the form[1]

$$Na + Cl = NaCl + 20{,}400 \text{ cal.} \tag{51}$$

We do not know the absolute amount of energy in a given mass of a substance but we have to deal here only with the differences, which appear as heat or as other forms of energy.

When heat is generated we speak of an exothermic, if it is absorbed, of an endothermic process, in which case the quantity of heat will have the negative sign.

According to the chemical reaction different terms are used, for example, heat of formation, heat of neutralization, heat of combustion, etc. The following examples will explain this:

(*a*) Heat of formation:

$$Zn + S + 4O = ZnSO_4 + 230{,}000 \text{ cal.} \tag{52}$$
$$Cu + S + 4O = CuSO_4 + 183{,}000 \text{ cal.} \tag{53}$$
$$2H + S + 4O = H_2SO_4 + 192{,}500 \text{ cal.} \tag{54}$$

(*b*) Heat of hydration:

$$Zn_2SO + 7H_2O = ZnSO_4,\ 7H_2O - 4{,}260 \text{ cal.} \tag{55}$$

(*c*) Heat of solution:

$$ZnSO_4 + aq = ZnSO_4,\ aq + 18{,}500 \text{ cal.} \tag{56}$$
$$CuSO_4 + aq = CuSO_4,\ aq + 15{,}500 \text{ cal.} \tag{57}$$
$$H_2SO_4 + aq = H_2SO_4,\ aq + 18{,}000 \text{ cal.} \tag{58}$$
$$NaCl + aq = NaCl,\ aq - 1{,}200 \text{ cal.} \tag{59}$$

(*d*) Heat of neutralization:

$$NaOH,\ aq + HCl,\ aq = NaCl,\ aq + 13{,}700 \text{ cal.} \tag{60}$$

(*e*) Heat of combustion:

$$C + 2O = CO_2 + 103{,}000 \text{ cal.} \tag{61}$$

[1] Frequently also the physical state is indicated by certain symbols, so as to show whether the substance is in the solid, liquid or gaseous state.

In some chemical reactions, for example, when zinc is dissolved in sulphuric acid, gases are produced, and thus external work is done in addition to the production of heat. The amount of energy thus set free should be included in the heat of reaction.

An important law for thermochemical measurements was discovered by Hess in 1840: "The total heat produced by a chemical change of a system *A* of substances to a system *B* is independent of the manner of transition from one to the other." This law of constant heat sums, which is only a special case of the law of conservation of energy (see § 318), allows the calculation of the heat of reaction, even in cases where the reaction is impossible.

As an application of Hess's law we will derive the following heats of reaction:

(1) By the addition of (52) and (56) we obtain

$$Zn + S + 4O + aq = ZnSO_4, aq + 248500 \text{ cal.} \qquad (62)$$

From (54) and (58)

$$2H + S + 4O + aq = H_2SO_4, aq + 210500 \text{ cal.} \qquad (63)$$

or

$$H_2SO_4, aq - S - 4O - aq = 2H - 210500 \text{ cal.} \qquad (64)$$

Adding (62) and (64)

$$Zn + H_2SO_4, aq = ZnSO_4, aq + 2H + 38000 \text{ cal.} \qquad (65)$$

(2) In a similar way we may derive

$$Cu + H_2SO_4, aq = CuSO_4, aq + 2H - 12000 \text{ cal.} \qquad (66)$$

(3) By combining the last two equations

$$Zn + CuSO_4, aq = Cu + ZnSO_4, aq + 50000 \text{ cal.} \qquad (67)$$

i. e., if a gram atom of zinc displaces a gram atom of copper from a copper sulphate solution 50000 calories are developed.

268. Heat of Combustion. From an engineering point of view a knowledge of the exact chemical reaction is not required. From this point of view the amount of heat evolved when one gram of the substance is burned is called its heat of combustion, the term being used here in a different sense from that in the last article.

HEAT OF COMBUSTION IN CAL. PER GRAM.

Substance.	H	Substance.	H
Hydrogen	34700	Alcohol	7183
Gunpowder	508 to 807	Illuminating gas	5200 to 6365
Dynamite	1290	Various woods	4100 to 4500
Sulphur	2220	Anthracite	7844

CONDUCTION OF HEAT.

269. Exchange of Heat between Bodies at Different Temperatures. If differences of temperature exist between the bodies of a system (or the parts of one body) there is always a tendency towards equalization of temperature by an exchange of heat between the bodies. Frequently three modes of transference of heat are distinguished.

a. Conduction. This is a transfer of heat from one portion of matter to another at lower temperature and in contact with it.

b. Convection. We have seen in § 257 that convection of fluids results in an equalization of temperature, though the actual transfer of heat from one portion of the system to another takes place only when the heated, moving particles come into contact with colder bodies and thus give up heat by conduction.

c. Radiation. Radiation is a transfer of energy from one body to another without heating of the intervening medium. Energy passes into space in such a manner that it can no longer be measured directly in heat units. We can say therefore that heat energy of the warmer body becomes a form of energy called energy of radiation, which passes out into space and upon being absorbed by another body raises the temperature of the latter, *i. e.*, is retransformed into heat energy. Though in the end there is an exchange of heat, radiation is not a direct transference of heat (see § 306).

Conduction is therefore the only process in which energy passes from one body to another or along a body in such a way that it can be measured as heat energy all along its path.

270. Conduction of Heat in Solids. Heat always travels from hotter to colder matter and the exchange is the more rapid the larger the difference of temperature per unit of path. The quantity $(t_2 - t_1)/l$ is called the *temperature gradient* in the direction in which l is measured.

The conductivity or power to transmit heat varies considerably with different substances. Thus a burning match can be held until the flame touches the fingers; one end of a piece of glass may be heated in a Bunsen flame and it will hardly become warm a few centimeters from the hot end, while a metal rod of the same size under similar conditions becomes too hot for the touch at a much greater distance from the flame.

Metals at a somewhat lower temperature than the human body feel cooler than poor conductors at the same temperature because they carry away the heat from the hand faster than it is supplied. If metals are at a higher temperature than the body they will appear hotter than poorer conductors of heat.

The difference in conductivity is easily demonstrated. Wrap a sheet of thin paper around a cylinder one end of which is made of metal and the other of wood. If the cylinder is then held in a flame the paper will char where it covers the wood, but it will not be changed where it is in contact with the metal. The line where the two substances join is clearly marked on the paper. The metal has carried off enough heat to protect the paper from burning.

In fact the flame in the neighborhood of a cool good conductor does not actually touch the conductor but is cooled far enough to prevent combustion. We can boil water in a paper tray because the temperature of the latter cannot rise above that of boiling water and consequently will not burn.

In Davy's safety lamp, used for the protection of miners from explosion of gases, a fine wire gauze surrounds the light and prevents the flame of the gas which may pass into the lamp and burn there, from igniting the gas outside. To show this effect place some wire gauze over a Bunsen burner and light the gas either above or below. The gas on the other side will not be affected until the separating gauze becomes red hot. Then the flame breaks through. Davy's lamp will therefore be a safeguard only as long as the burning gas does not overheat the wire gauze.

The conductivity of different metals varies widely. If two rods of the same size, but of different materials, are covered with wax and one end of each is kept in a flame for some time, the wax will have melted to a farther distance from the flame on the rod having the greater conductivity. At first, *i. e.*, before the steady state is reached, the heat capacities of the substances greatly influence the result and therefore no definite conclusions can be drawn from this experiment, while the temperatures along the bars still vary with the time. After the flow of heat has become steady the temperature at every point remains constant and there is a gradual decrease in temperature from the hotter end to the

other, the temperature slope being the steeper the smaller the conductivity.

271. Coefficient of Thermal Conductivity. If two parallel plane surfaces of a body, a distance l apart, are kept at the constant temperatures t_2 and t_1 the quantity of heat transferred through any cross-section $A = a \times b$ at right angles to the flow of heat is proportional to the cross-section, to the temperature gradient $(t_2 - t_1)/l$ and to the time τ during which the process continues.

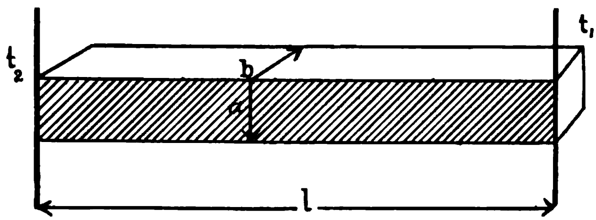


FIG. 147.

The proportionality factor, k, which depends upon the nature of the conducting substance, is called the coefficient of thermal conductivity of that substance.

Writing this in the form of an equation we have

$$H = kA\tau \frac{t_2 - t_1}{l}. \tag{68}$$

The coefficient of thermal conductivity of a substance is therefore the *time rate with which heat passes through unit area if the temperature gradient is unity.* The quantity of heat is usually expressed in calories, the temperature in degrees and the rest of the quantities entering the equation in C. G. S. units.

Frequently the relative conductivity is given referred to that of silver (or of copper) as 100, or, in the case of liquids and gases, referred to water and air respectively.

The experimental determination of this physical constant presents many difficulties, mainly on account of the loss or gain of heat by the bounding surfaces of the conducting substances and the disturbance produced by the introduction of devices for measuring the temperature gradient. The results obtained by different experimenters vary considerably from each other, and the values

in the following table must be considered as only approximately correct. Conductivity plays an important rôle in the heating of boilers. It is not alone the metal of the boiler that must be considered in this respect but also the soot or grease on the outside and the scale deposited from the water on the inside. In engineering practice it is assumed that 0.01 in. of soot or grease, 0.1 in. of scale and 10 in. of metal are equivalent to each other.

Thermal and electrical conductivity seem to be closely related phenomena. The ratio between the coefficient of thermal conductivity and that of electrical conductivity is nearly the same for many metals. Both these quantities decrease with the temperature and, in the case of pure metals, their temperature coefficients are the same.

The electromagnetic theory of light shows that a close connection exists between the optical constants of a substance and its electrical conductivity. It seems therefore not improbable that heat conduction may, at least in part, be a radiation phenomenon between the molecules.

272. Conductivity of Liquids and Gases. Liquids have in general only a low conductivity for heat as compared with solids. To show this the liquids must be heated from above in order to prevent convection. We can thus boil water in the upper part of a test-tube without materially changing the temperature of the lower layers.

COEFFICIENTS OF THERMAL CONDUCTIVITY.

(k calories pass in one second through 1 cm^2. of a plate 1 cm. thick, if the difference of temperature on the two sides is 1° C.)

Substance.	k	Substance.	k
Silver	1.00	Glass	0.0015
Copper	0.90	Ice	0.005
Gold	0.70	Basalt	0.005
Aluminium	0.50	Fresh snow	0.0001
Zinc	0.25	Pine wood ‖ fiber	0.0003
Iron	0.15	Pine wood ⊥ fiber	0.0001
Bismuth	0.02	Felt	0.00009
Mercury	0.016	Flannel	0.00004
Water	0.0015	Hydrogen	0.00032
Alcohol	0.0004	Air	0.00005
Benzol	0.0003	Carbon dioxide	0.00003

Gases are still worse conductors. The poor conductive qualities of solids of loose structure, as wood, paper, sawdust, wool, asbestos, etc., is mainly due to the air confined in the interstices between

the solid material. In such substances the transference of heat must be very slow. Upon this, for example, depends the "warmth" of woolen clothes.

The coefficient of thermal conductivity of gases is independent of the pressure, except when the pressure is very small.

CHANGE OF STATE.

273. Changes of State with Rise of Temperature. When a solid crystalline body is heated its temperature rises until it begins to melt or fuse. If the heat is not supplied too fast, so that the temperature is uniform throughout, no further change in temperature will be observed as long as a part of the body is still in the solid state. After the whole mass is melted the temperature of the liquid rises upon continued heating to a point at which boiling sets in. Then the temperature remains constant again until the substance is entirely changed to vapor, when the temperature once more begins to rise. At the stationary temperatures which, under the existing atmospheric pressure, are characteristic constants for each substance, the solid and the liquid—or the liquid and the vapor—are in equilibrium with each other, so that a mixture of the two will remain unchanged if no heat is supplied or removed.

While we shall consider in the following only the solid, liquid and gaseous states and the phenomena connected with a transition from one of these into another (including the phenomena of solution), the change of one allotropic modification of a substance to another belongs clearly in this chapter. We find here also definite transition temperatures, absorption or evolution of heat during the transformation and a change of density and other physical properties. Rhombic sulphur, for example, becomes under atmospheric pressure monoclinic at 95.5° and during this process 2.53 calories are absorbed by one gram of sulphur.

In considering changes of state from this more general point of view the term "phase" is widely used to denote any component of a system of bodies, which is homogeneous and differs in its physical properties from the rest so that it can be removed from the system by purely mechanical means. A mixture of ice and water consists of two phases, a solution containing undissolved salt and in contact with an atmosphere of water vapor is a system of three phases, to which one more phase must be added if also ice be present. From this point of view allotropic modifications are separate phases. The phenomena studied in this chapter are only special cases of the general equilibrium between any number of different phases.

274. Melting Point. The melting point or fusing point of a substance *is the temperature at which the solid and liquid states are in equilibrium* under the existing pressure. It is usually referred to atmospheric pressure. Above this temperature the substance is a liquid, below it, as a rule, a solid. (We disregard for the present the possible formation of vapor on the surfaces of the solid and liquid. See § 294). When a liquid is cooled solidification sets in when the melting point is reached and for this reason the melting point is frequently also called the freezing point. (For supercooling see § 275.)

MELTING POINTS AND BOILING POINTS UNDER ATMOSPHERIC PRESSURE.

Substance.	Melting Point.	Substance.	Melting Point.	Boiling Point.
Aluminium	657°	Cadmium	321°	778°
Gold	1063.5	Copper	1084	about 2100
Iridium	2200	Lead	327	1170
Iron	1100	Mercury	—38.8	357
Steel	1350	Phosphorus	44.3	287
Nickel	1484	Sodium	97.5	about 800
Osmium	2500	Sulphur	115	444.5
Platinum	1778	Tin	232	about 1500
Silver	961	Zinc	419	" 930

(see also table on p. 246.)

275. Supercooling. A liquid can generally be cooled below its normal fusing point without solidification. Thus water, if protected from mechanical disturbance or if covered with a layer of oil, will remain liquid at — 10° or in capillary tubes even at — 20°. This phenomenon is called "supercooling," "undercooling" or "surfusion." Supercooled liquids are in unstable equilibrium; for, if disturbed, they will immediately pass into the stable state, as a mixture of solid and liquid, at the normal fusing point.

The rise of temperature accompanying this readjustment is due to the large amount of heat set free by crystallization (see § 279). Hyposulphite of sodium, $Na_2S_2O_3 + 5H_2O$, after being melted at a temperature above 50° may be cooled and kept liquid at ordinary temperatures. But as soon as a crystal of the salt is added solidification begins and the temperature rises to the melting point 47.9°. (Other examples are glaubersalt, $Na_2SO_4 + 10H_2O$ with a melting point at 32.38°, and sulphur which melts at 115°.)

Since the temperatures reached when solidification under these conditions sets in, *i. e.,* the melting points, are constant for a given substance, such points may be used to calibrate a thermometer.

Amorphous bodies, as wax, glass, etc., have no definite melting point, but the solid when heated becomes more and more plastic; there is no sudden transition from the solid to the liquid state at a definite temperature. On the other hand, when originally liquid, these substances become on cooling more and more viscous and finally very hard; but there is no specific temperature at which we can speak of a transition from the liquid to the solid state.

After a liquid has been cooled so far that it has become solid without passing through a definite freezing point the resulting amorphous solid is more closely related to the state at a higher temperature, at which it would be called a liquid, than it is to the solid which has formed by crystallization from this liquid. Some substances may exist as well in the amorphous as in the crystalline solid state, for example, quartz as quartz crystals and quartz glass. The latter is obtained by fusing quartz crystals at about 1,500°; the mass remains amorphous when cooled to ordinary temperatures, and has quite different properties from the crystals.

276. Solutions and Alloys. In dilute solutions the freezing point is lowered below that of the pure solvent and the lowering is nearly proportional to the concentration. Moreover only the pure solvent crystallizes out and thus the solution becomes more concentrated, accompanied by a further lowering of the freezing point.

A curve giving the "freezing points" of dilute solutions as a function of the concentration may be considered as the saturation curve of the solution with respect to the solvent; it marks the equilibrium between the solution and the solid solvent.

On the other hand if we cool a warm concentrated solution the dissolved salt will begin to crystallize out at a definite temperature, the solution will become more dilute, more salt will crystallize out on further cooling and so forth.

The curve representing the relation between the temperature and the concentraton of the solution when saturated with respect to the salt is called the solubility curve. In this case the temperature decreases with decreasing concentration. The two saturation curves are represented in Fig. 148, where they are drawn as

straight lines. They meet at the point *P*. If this point is reached from either direction, *i. e.*, by continuous cooling of either a dilute or a concentrated solution the whole mass solidifies as a mechanical mixture of the two components in the solid state.

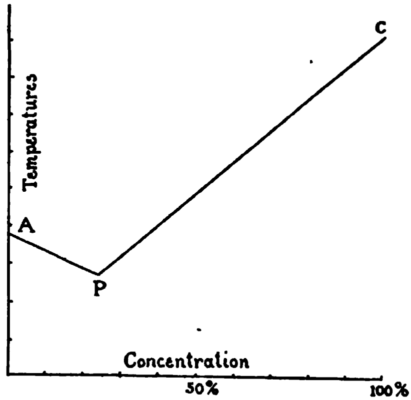


FIG. 148.

This mixture which has (under a given pressure) a perfectly definite composition, is called a *cryohydrate*. The cryohydrate for a sodium chloride solution contains 26.6 per cent. NaCl and solidifies at —23°. (NaCl melts at about 800°).

Alloys and mixtures of crystals behave similarly. Thus we find in general that an alloy has a lower melting point than the pure metals which it contains; and if the curves are plotted showing the dependence of the melting point upon the amount of the metals present curves similar to those in the case of solutions would be obtained. The alloy corresponding to what was called in solutions the cryo-hydrate is called a "*eutectic*" mixture. Rose's metal, an alloy of bismuth, lead and tin, melts at 96°, Wood's metal which contains also some cadmium at 70°.

Usually the behavior of alloys is complicated by the fact that neither the pure solvent nor the pure dissolved body crystallize out, but definite mixtures or solid solutions, which in general have a different concentration from that of the remaining liquid. In such cases the melting and freezing of an alloy of definite concentration begin at different temperatures. The existence of chemical compounds with distinct freezing points may complicate the curve still further.

See Landolt and Boernstein's Tables, 5th ed., p. 112a–112m.

277. Change of Volume during Fusion. Most substances expand during fusion and contract during solidification, so that, if both states are present, the solid sinks in the liquid. Notable exceptions to this rule are water, bismuth, cast iron and type metal.

The density of ice is only 0.916 gram/cm.^{3} at 0° while that of the liquid state is, at the same temperature, 0.99987 gram/cm.^{3}

This expansion during freezing is of great importance in nature, preventing a complete freezing of lakes during the winter and accelerating the disintegration of rocks whose fissures are filled with water. If water is enclosed in a vessel the pressure exerted during freezing is very large, and will easily burst thick metal tubes (freezing of water pipes).

The expansion of iron and type metal during solidification is used for an exact reproduction of the mold into which they are cast.

278. Influence of Pressure Upon the Melting Point. The melting point depends to a slight extent upon the pressure. Substances, such as water, which expand upon solidification have their melting points lowered by pressure and *vice versa.*

The freezing point of water is lowered 0.0072° by an increase in pressure of one atmosphere *i. e.,* it requires a pressure of nearly 150 atmospheres to prevent the freezing of water at —1° (see also § 328).

Snow at 0° is easily formed into a snowball by the pressure of the hand; if the pressure is released the water in the interior will freeze again. But in colder weather the snow will not "pack." Two pieces of ice pressed against each other in warm water will freeze together as soon as the pressure is removed. This process is called *regelation.* A very instructive experiment is the following, first performed by Bottomley. Place a block of ice with its ends upon supports and hang over it a heavy weight by means of a fine wire. The pressure on the under side of the wire will melt the ice and the wire will slowly cut through it. Above the wire the pressure has the normal value and, since heat is absorbed during fusion, the wire is at a temperature below 0° and the water will freeze again above the wire so that a solid piece of ice is formed again. If an attempt is made afterwards to break the ice, the break is as liable to occur at any other place as in the plane through which the wire passed.

The movement of glaciers is partly due to regelation. The load resting upon the lower parts is large enough to melt the ice wherever it is in contact with the resisting rock. The structure

of the ice, which is an aggregation of many fragments oriented in different ways, facilitates the motion.

279. Heat of Fusion. We have seen (§ 273) that heat is required to melt a crystalline solid but that the temperature remains constant during this process. The larger part of this heat energy serves to do the internal work necessary to produce the change of state, while the external work is usually small. When a liquid solidifies the same amount of heat is set free as is absorbed during liquefaction.

Heat of fusion *of a body* (or latent heat of fusion) is the quantity of heat necessary to melt the body without changing its temperature. Since this quantity of heat is proportional to the mass of the body we have

$$H = LM.$$

The proportionality factor L, though not heat, is called the heat of fusion *of the substance* and may be defined as the ratio of the *quantity of heat which is required to change the body at the melting point from the solid to the liquid state to its mass,* and is numerically equal to the quantity of heat necessary to melt unit mass of the substance. Its unit is a calorie per gram.

This heat may be measured by the method of mixtures. Let a solid of mass M_1 at the temperature t_1 be introduced into a calorimeter whose total water equivalent (calorimeter, stirrer, thermometer and water) is M_2 and temperature t_2. The latter must be high enough to ensure a temperature of the mixture, t, above the normal melting point, t_0, of the solid. Let the heat capacities of the substance investigated be c_1 in the solid and c_2 in the liquid state.

By the principle of equal heat exchanges

$$M_1[c_1(t_0 - t_1) + L + c_2(t - t_0)] = M_2 c_w (t_2 - t) \qquad (69)$$

from which L may be calculated.

It is apparent that the reverse process may be applied, *i. e.*, the substance may be solidified in the calorimeter.

The heat of fusion of water is 80 calories per gram. This large value has some influence upon climatic conditions in preventing a sudden transition from fall to winter, and winter to spring.

The heat of fusion of water being known equation 69 may be used for the determination of the specific heat of a substance, by surrounding it by ice at 0° while the body of mass M_2 cools from the temperature $t°$ to 0°, and measuring the amount M_1 of the ice melted. In this case, equation (69) simplifies to

$$M_1L = M_2c_2t_2 \tag{70}$$

280. Freezing Mixtures. During the process of solution heat is frequently absorbed, especially in case of some nitrates and chlorides (see § 267, equation 59), and this may be employed for the lowering of temperatures. The effect will be much more marked if heat of fusion disappears at the same time.

A mixture of snow or ice with common salt is universally employed as a freezing mixture, resulting in the formation of a concentrated salt solution. The best results will be obtained if the quantities are chosen in a proportion corresponding to the cryohydrate. The temperature reached is then the melting point of the cryohydrate. Thus with a mixture of ice and salt —23° may be reached, with one of ice and calcium chloride even —42°.

281. Vaporization. Vaporization is the transition of a liquid or a solid into the gaseous state. The change from a liquid to a gas is called evaporation and takes place continuously at the free surface of a liquid, *i. e.,* a surface which is not in direct contact with a solid or another liquid. This is easily explained by the molecular theory.

During evaporation some molecules of the liquid have a velocity large enough to carry them through the surface beyond the sphere of influence of the molecules forming the bulk of the liquid. They become thus gaseous molecules and as such exert, if enclosed in a vessel, a definite gas pressure due to their impact against the walls.

In ordinary language the term "vapor" is used for a gas if it is thought of as being formed from a liquid or a solid (but see also § 301).

In a manner similar to that described above a solid may change into vapor without previous fusion. The passage from the solid directly into the gaseous state is called sublimation (§ 293).

The reverse process, namely the change of a gas or vapor into a liquid or solid, is called condensation.

282. Evaporation in a Closed Vessel. If several cubic centimeters of a liquid are introduced into a Torricellean vacuum (see § 218) a part of it will evaporate and due to the pressure of the vapor the mercury column will fall, but soon all visible evaporation will cease and an equilibrium be established, since as many molecules leave the remaining liquid in unit time as enter it from the vapor.

FIG. 149.

The vapor which is in equilibrium with its liquid is called "saturated" vapor. At a given temperature the physical properties of saturated vapor, as density, pressure, etc., are perfectly defined.

The difference between the atmospheric pressure and that corresponding to the height of the mercury in the tube containing the liquid with its vapor is thus a constant for a given substance and given temperature. It is the maximum pressure the vapor can have at that temperature and is called the *vapor tension.*

The volume occupied by the vapor does not affect the vapor tension, as is easily shown by using tubes of different width or length for the experiment (Fig. 149). Upon compression the vapor will condense and, if the pressure is constantly kept higher than the vapor tension, finally all vapor will disappear.

If only a very small amount of the liquid were introduced in the tube, all of it would evaporate and the resulting vapor be unsaturated or superheated. If such vapor is placed in contact with the liquid, further evaporation will take place, until the vapor is saturated.

283. Vapor Tension Dependent upon Temperature. The vapor tension of a substance may be measured as described in the last article or by any other method used for measuring gas pressures. It is a definite function of the temperature and increases somewhat faster than the latter. The vapor tension curve for water is

drawn in Fig. 150 and the numerical values are given in the table below.

VAPOR TENSION OF WATER IN CM. OF MERCURY.

Degrees.	Tension	Degrees.	Tension.	Degrees.	Tension.
—20	0.096	40	5.50	100	76.0
—10	0.216	50	9.22	110	107.5
0	0.458	60	14.92	150	357
+10	0.918	70	23.38	200	1165
20	1.741	80	35.55	250	2980
30	3.156	90	52.60	300	6760

At any point of the curve water and water vapor are in equilibrium, but not at any other point of the diagram. Suppose water and its vapor to be in a closed vessel (barometric tube for example) allowing a variation of temperature and pressure. At 20° the pressure of the vapor is 1.74 cm. of mercury. Let the pressure be increased say to 7.15 cm. of mercury; the vapor will all condense and no vapor can exist under this pressure, except when the temperature is raised to at least 45°. On the other hand if the pressure is reduced and kept low, evaporation will continue until finally all the liquid is changed to vapor.

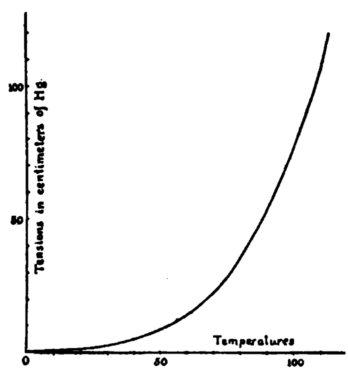


FIG. 150.

284. Dalton's Law. If the space over a liquid, instead of being a vacuum, contains another gas which does not act chemically upon the vapor, evaporation takes place, but at a diminished rate, until the pressure of the vapor alone equals the vapor tension. The presence of other gases has no influence upon the total amount of evaporation. In the end there is a mixture of gases each of which might be considered as having its own individual pressure. The pressure of the mixture equals the sum of the pressures of its constituents.

It is therefore a common experience that water evaporates freely in air at ordinary temperatures except on very moist days. The

total air pressure is considerably above the vapor tension of water, but the amount of water vapor is usually so small that its partial pressure is below it. "Vapor laden" air is a mixture of air and water vapor. Since the latter is less dense than air, it is evident that the mixture weighs less than the same volume of dry air under the same pressure, some of the air having been displaced by the vapor. For this reason the atmospheric pressure decreases when there is much moisture in the air.

The larger the difference between the vapor tension and the partial pressure of the vapor above the liquid, the greater is the rate of evaporation. For this reason evaporation is accelerated by blowing relatively dry air over the surface of the liquid (wind effect) or by a rise of temperature. Naturally under otherwise equal conditions an increase of the surface will also increase the amount evaporated per second.

285. Ebullition. Boiling Point. If the vapor tension of a liquid becomes as great as or slightly greater than the pressure of the atmosphere (or mixture of gases) in contact with it, evaporation will take place not only at the free surface but also below it, usually on the walls of the vessel containing the liquid, *i. e.*, vapor bubbles will be formed in the interior of the liquid and rise to the surface. This phenomenon is called boiling or ebullition.

The boiling point of a liquid is therefore *the temperature at which its vapor tension equals the pressure of the surrounding atmosphere.* The normal boiling point is always referred to an atmospheric pressure corresponding to a barometric height of 76 cms. at 0°.

BOILING POINTS OF ORGANIC LIQUIDS UNDER ATMOSPHERIC PRESSURE.

SEE ALSO TABLE ON P. 219.

Substance.	Boiling Point.	Substance.	Boiling Point.
Alcohol	78°	Ether	35°
Anilin	184	Glycerin	290
Chloroform	61	Toluene	110

286. Influence of Pressure upon the Boiling Point. From the definition of the boiling point and from the variation of vapor tension of a liquid with temperature, it follows that the vapor ten-

sion curve also represents the dependence of the boiling point upon the external pressure and that the boiling point is lowered by a decrease in pressure and vice versa.

After water has been made to boil in an open round-bottomed flask so as to expel the air, cork the flask while removing it from the flame, invert it and place a moist cloth on the bottom of the flask. Some of the vapor in the flask will condense, the pressure be reduced and vigorous boiling take place at the reduced pressure, though the temperature of the water may have fallen considerably.

Or place a flask with water heated to about 50° or 60° under the receiver of an air pump and remove the air and vapor. The water will begin to boil as soon as the pressure is sufficiently reduced.

BOILING POINT OF WATER UNDER DIFFERENT PRESSURES (IN CM. OF HG).
SEE ALSO TABLE ON P. 226.

Pressure.	Boiling Point.	Pressure.	Boiling Point.	Pressure.	Boiling Point.
73.0	98°.88	75.0	99°.63	77.0	100°.37
73.5	99.07	75.5	99.82	77.5	100.55
74.0	99.26	76.0	100.00	78.0	100.73
74.5	99.44	76.5	100.18	78.5	100.91

The dependence of the boiling point upon the pressure is much more marked than that of the melting point; for example, a change in pressure from atmospheric pressure corresponding to 1 cm. of mercury changes the boiling point of water 0.37°.

287. Superheating. Just as a liquid can be cooled below the freezing point without solidification as long as the solid state is absent, so also vaporization may be prevented if care is taken to exclude the gaseous state, *i. e.*, to let the liquid have no free surface. If a clean glass tube about 90 centimeters long and closed at one end is filled with mercury and a little air-free ether and then inverted in a deep mercury trough, the ether will rise through the mercury to the top. If no air space is left, no evaporation will take place, though the ether may actually be under tension (when the mercury column is longer than 76 cm.). A slight blow against the tube will bring about sudden evaporation and the establishment of stable equilibrium.

Boiling of a liquid may be retarded in a similar manner. A spherule of water suspended in oil of the same density may be heated to 150° without evaporation taking place. A liquid in which boiling is thus prevented at a temperature higher than the boiling point is said to be superheated.

Usually minute air bubbles adhere to the walls of a vessel containing a liquid and at these places the formation of the vapor bubbles begins when the temperature approaches the boiling point. This will usually occur at the place where the heat is applied.

After the air has been completely removed by continued boiling the liquid can be heated several degrees above the normal boiling point before vapor is formed in the interior; then boiling sets in with explosive violence. To prevent this "bumping" porous or rough substances are placed in the bottom of the vessel or a fresh supply of the liquid containing air is added.

Since the liquid may be at a somewhat higher temperature during boiling than its normal boiling point, determinations of the latter are made with the thermometer bulb kept slightly above the surface and surrounded by the saturated vapor only.

The intermittent action of geysers is also explained by the overheating of the lower portions of the water in the geyser tube.

288. Boiling Point of Solutions. From solutions which contain a non-volatile substance only the solvent evaporates, while the dissolved salt crystallizes out as the solution becomes more concentrated. A solution has always a lower vapor tension at a given temperature than the solvent and the lowering is approximately proportional to the concentration.

Consequently the vapor tension curve of a solution lies below that of the solvent and the temperature at which boiling takes place must be higher than that for the pure solvent. Thus 35.5 parts of $NaCl$ dissolved in 100 parts of water lowers the vapor tension at 100° almost 18 cm. of mercury and the boiling point is raised to 107.5°.

The temperature of the vapor just above boiling solutions equals the boiling point of the solution and not the boiling point of the solvent, but a thermometer held in the vapor will in general indicate the latter temperature owing to condensation of the solvent on the bulb.

The relations are much more complicated for solutions in which the dissolved body also possesses an appreciable vapor pressure, for example, alcohol and water.

289. The Spheroidal State. If water is carefully placed by means of a pipette upon a red hot metallic plate it will spread out into a flattened spherule and evaporate quietly. This is called Leidenfrost's phenomenon and the water is said to be in the spheroidal state. The temperature of the water while in the spheroidal state is always a few degrees below the boiling point.

The liquid is not in contact with the plate but shielded from it by a cushion of vapor. This may be proved in the following way. Connect the two terminals of an electric circuit containing a bell to the plate and the water-drop respectively. The bell does not ring since the circuit is broken between the drop and the plate.

Remove the flame from below the plate and let it cool. At a certain temperature the drop will come into contact with the hot plate as proved by a ringing of the bell, and violent boiling sets in.

This phenomenon can be observed with many other liquids whose boiling points are considerably below the temperature of the bodies with which they come into contact.

290. Change of Volume during Evaporation. In general the change of volume is considerable during evaporation. But it is a noteworthy fact that the higher the temperature rises the more do the densities of the liquid and its saturated vapor approach each other.

The results for water are shown in the following table:

DENSITY OF WATER AND SATURATED WATER VAPOR IN GRAMS./CM.^{3}

Temperature.	0°	50°	100°	150°	200°
Liquid	0.999	0.988	0.959	0.917	0.86
Vapor	0.00000475	0.000083	0.000606	0.0026	0.0079

By following the vapor tension curve to higher temperatures we will thus finally arrive at a point where the densities of the liquid and gaseous states are the same and at this point the two states must become physically identical, *i. e.*, the vapor tension curve must end at this point.

Fig. 151 shows Ansdell's results for hydrochloric acid, HCl. The volumes occupied by 6.8 milligrams of the substance are plotted as functions of the temperature.

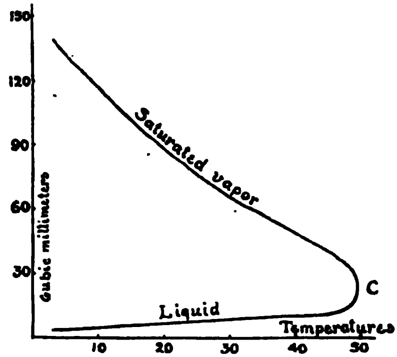


FIG. 151.

The vapor pressure curve must end at about 50°, since at the point *C* the curves for the two volumes join. We shall see later that this point is identical with the critical point (§ 301).

291. Heat of Vaporization. It is a well known fact that heat must be supplied continuously to a liquid in order to keep it boiling. If the rate of application of heat is increased the boiling becomes more violent, *i. e.*, the rate of evaporation increases; but the temperature is not changed appreciably. Evidently the heat added to the liquid is needed to transform the substance from the liquid to the gaseous state. Quiet evaporation from the surface of a liquid at a temperature below the boiling point is also accompanied by an absorption of heat. Thus a little ether dropped upon the skin causes a sensation of cold because the heat necessary for the evaporation of the ether is supplied by the skin.

Solids also require heat for sublimation, but since the amount of a solid evaporated under ordinary conditions is quite small, this is more difficult to prove experimentally.

Measurements have shown that the amount of any solid or liquid vaporized is proportional to the heat supplied. *Heat of vaporization of a substance is the quantity of heat necessary to change the solid or liquid to saturated vapor divided by the mass,* and is numerically equal to the quantity of heat necessary to evaporate unit mass. In the case of liquids it is as a rule referred to the normal boiling point.

This definition corresponds to that of heat of fusion. The heat energy added to the substance serves in part to increase the internal molecular energy and in part to do the external work of expansion which accompanies the process. Upon condensation the same

amount of heat is liberated as is absorbed during vaporization. This can be shown easily by leading steam into a vessel of cold water. The steam condenses, gives out heat and heats the water in the vessel. Let M_1, c_1 and t_1 be the mass, heat capacity and temperature of the vapor, t_0 its boiling point and c_2 the heat capacity of the liquid. According to the principle of equal heat exchanges we obtain an equation similar to (69)

$$M_1[c_1(t_1 - t_0) + L + c_2(t_0 - t)] = M_2 c_w (t - t_2). \qquad (72)$$

From this L can easily be determined.

The heat of vaporization depends greatly upon the temperature at which vaporization takes place. Griffiths derived from his own experiments and those of Regnault and Dieterici the following formula for the heat of vaporization of water between 0° and 100°:

$$L = 596.63 - 0.601t \text{ cal. per gram}, \qquad (71)$$

where t denotes the temperature. Under normal conditions the heat of vaporization of water is therefore 536.5 cal. per gram. More recent investigations have shown that this value is probably too small and ought to be 538.

If this formula were quite general the heat of vaporization would become zero at 990°. The critical point for water is however 365° and it is probable that at this point the heat of vaporization disappears.

292. Cold by Evaporation. The heat of vaporization must be supplied by the liquid itself and by the bodies surrounding it; they will therefore be cooled. A vessel filled with water and standing for some time undisturbed in a room has a lower temperature than the air, especially if the evaporating surface is large, which is the case if the liquid is kept in a porous vessel. By increasing the rate of evaporation the effect is naturally intensified.

Water can even be frozen if rapid evaporation is kept up. This may be done by placing it under the receiver of an air pump and removing the vapor as fast as it forms. Of course care must be taken to avoid good thermal contact with the metallic parts of the machine. Use a watchglass or paper tray for holding the water and place it on corks, or, instead of using a separate dish, make a shallow hole in a large cork. To accelerate the action a dish with concentrated sulphuric acid should be placed under the belljar.

Water can thus be frozen in two to three minutes, sometimes even while it is boiling at the low pressure.

In commercial freezing plants an air-tight vessel, the "freezer," contains liquid ammonia. The ammonia vapor is by a pipe placed in communication with a vessel containing water. The latter absorbs ammonia freely, thus reducing the vapor pressure over the ammonia and increasing its rate of evaporation. The temperature of the ammonia is greatly reduced and water can easily be frozen in vessels surrounding the "freezer." When the water which has absorbed the ammonia is fairly saturated, it is pumped into a boiler in which the ammonia is expelled from the solution, to be condensed and returned again to the freezer. The process is thus made continuous.

If the valve of a liquid carbon dioxide tank is opened, the small stream issuing from it evaporates so rapidly that it freezes in the form of a fine snow which has a temperature of —79.2°.

293. Sublimation. Not only liquids but also solids evaporate though at a much slower rate than liquids. It is well known that a piece of ice diminishes in size even if constantly kept below the freezing point. Its vapor tension at 0° is 4.58 mm., but decreases rapidly with the temperature, being only 1.97 mm. at —10° and 0.79 mm. at —20°.

The change from the solid to the gaseous state is called *sublimation,* though frequently the reverse process is included in the meaning of the term. At ordinary pressures and temperatures camphor, arsenic and many less familiar substances sublime. Solid carbon dioxide volatilizes at —79.2° under atmospheric pressure, without passing through the liquid state.

With most solid substances the process is so slow at ordinary temperatures that it cannot be detected. Our sense of smell may give us in certain cases an indirect proof of the existence of the gaseous state.

The curve of the vapor tension of a solid as a function of the temperature is called the sublimation curve.

The heat of sublimation is evidently only a special case of the heat of vaporization and is larger than the heat of vaporization of the liquid state. At the triple point (see next article) it is equal to the sum of the heat of fusion and the heat of vaporization of the liquid.

294. The Triple Point. In the preceding articles we have studied the conditions of equilibrium for the transition of one of the three states, solid, liquid and gaseous, into another and have seen that for each of the possible combinations an equilibrium exists only at one definite temperature under a definite pressure or vapor tension. This is represented by the three curves for fusion, evapora-

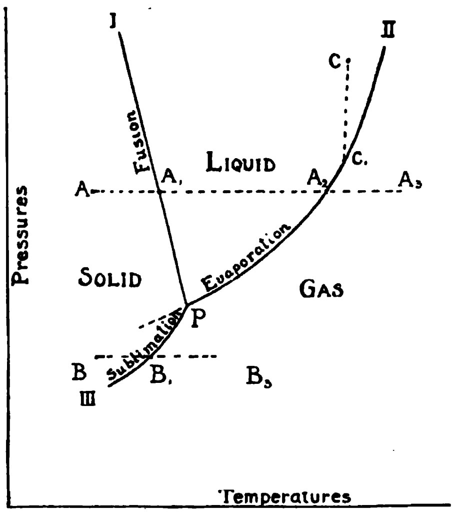


FIG. 152.

tion and sublimation. The common point of these three curves is called the "triple point" and only at this point can all three states exist in equilibrium. In Fig. 152 the three curves are drawn for a substance in which, as in ice, the melting point is lowered by increased pressure. For the water curve *I* would be very nearly parallel to the pressure axis since the depression of the freezing point is very small for this substance (see § 278).

The plane of the diagram is divided into three portions corresponding to the three states. From the figure the state of equilibrium of the substance under given pressure and temperature can be read off directly.

The triple point for water has the coördinates: $t = +0.0076°$ and $p = 4.6$ mm. of mercury. If a piece of ice, under a constant pressure, p, of more than 4.6 mm. of mercury and not in contact

with a gas (for example in an inverted tube filled with mercury) be heated from a low temperature, t, the change taking place will follow the straight line AA_1 parallel to the temperature axis. As soon as A_1 is reached melting will begin and the temperature will remain constant until the whole mass is melted. The line A_1A_2 corresponds to the heating of the liquid and at A_2 evaporation begins. If a mixture of ice and water kept at 0° is in contact with the atmosphere neither ice nor water is in equilibrium with the water vapor; if the mixture is kept in a closed vessel, filling it only partly the water would evaporate and the vapor condense to ice until finally only ice and vapor remain.

But if the solid is under a pressure smaller than that corresponding to the triple point, for example at a point B, then upon increase of temperature no melting will occur at B_1, but sublimation.

For camphor the pressure of the triple point is larger than one atmosphere and for this reason it cannot be melted in an open tube but sublimes upon being heated.

For carbon dioxide the triple point lies at $t = -56.24°$ and $p = 5.1$ atmospheres. Liquid carbon dioxide is kept in tanks under very high pressure. If this is suddenly reduced to one atmosphere by opening the valve, a change takes place which, neglecting the accompanying lowering of the temperature, may be represented by CC_1 in Fig. 152. Boiling will begin as soon as C_1 is reached, but this point lies still considerably above one atmosphere and the liquid and vapor cannot possibly coexist in equilibrium. Consequently boiling continues and owing to the absorption of heat of vaporization the temperature of the liquid is constantly reduced and the system moves along the line C_1P. From P on it must follow along PB_1, *i. e.*, it will solidify and finally reach along this line a point at which the solid and its vapor are in equilibrium under the existing vapor tension. We have seen that the temperature of this point is $-79.2°$. A boiling point for liquid carbon dioxide under atmospheric pressure does not exist.

Solid carbon dioxide can only be liquefied by increasing the pressure sufficiently. It may for example be placed in a strong glass tube and the latter sealed. The vapor produced will soon increase the pressure to more than 5.1 atmospheres, when liquefaction will take place.

295. Graphical Representation of the State of a Body. For an exact definition of the physical condition of a body a knowledge of the values of all its variable properties is required. For the present only three of these need to be considered, temperature, pressure and the volume occupied by unit mass of the substance. These are, however, not independent of each other, but connected by a definite relation, called the "equation of state," which in the simple case of perfect gases takes the form of the gas law (§ 256), equation (39).

The state of a substance is in general defined by only two of the three

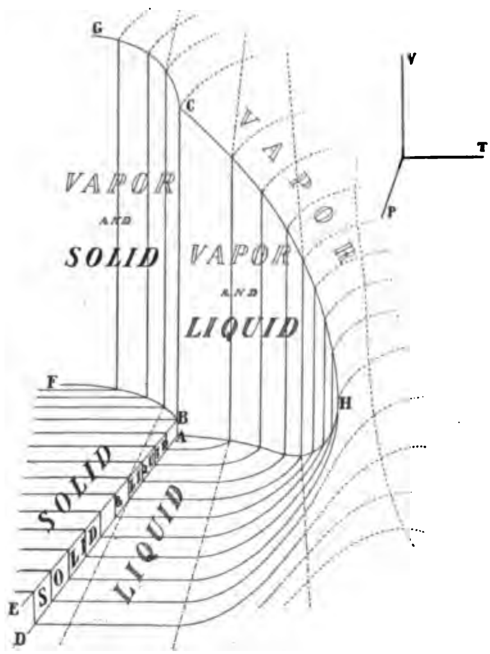


FIG. 153.

characteristic properties and can be represented by a plane diagram, with these two variables as coördinates. Which two are selected remains arbitrary. Thus in the discussion of evaporation temperature and pressure were used (Figs. 150 and 152), in another case, temperature and volume (Fig. 151).

The value of the third of the three properties is then perfectly defined for each point of the diagram except where, as at the evaporation curve, both variables remain constant, while the third property changes. The

variation of the volume during a change of state can therefore not be shown on the temperature pressure diagram.

The pressure-volume diagram is free from this objection and is therefore most frequently employed.

The plane diagrams mentioned are in fact nothing but the three projections of the surface, defined by the equation of state, upon one of the three coōrdinate planes containing either the p and v, or t and p, or t and v axes.

Fig. 153 is taken from a photograph of a model which serves to demonstrate the general appearance of the various surfaces representing the different states of a body which, like water, contracts upon melting. The lines from left to right in the figure are lines of constant pressure while the lines from front to back are constant temperature lines. The three surfaces *EBAD, GCBF* and *CHAB* represent the conditions under which two of the states may be in equilibrium. In each of these three cases the pressure remains constant, while a transformation from one state to another takes place at constant temperature. Therefore all the lines of constant pressure and temperature on these surfaces must be vertical to the pressure-temperature plane and the projections of the three surfaces upon this plane appear simply as curves: surface *EBAD* as the curve of fusion, *CHAB* as the vapor-tension curve and *GCBF* as the sublimation curve. The line *ABC*, which represents the conditions under which all three states may be in equilibrium, will appear as a point, the triple point. Such a projection was given in figure 152, the observer looking in the direction of the positive volume axis.

It should, however, be clearly understood that our choice is not restricted to these three variables, but that any other physical quantity better suited for the problem under discussion may be taken for a coōrdinate, for example, *PV*, if the variation of this quantity with the pressure is to be investigated.

296. Condensation of Water in the Atmosphere. If air containing moisture is cooled, a temperature will finally be reached at which the actual pressure of the water vapor equals the vapor tension and at this point condensation will begin, either to liquid water or, if the temperature is below 0°, to ice (hoar frost). This temperature at which the water vapor becomes saturated is called the *dew point* and depends evidently upon the amount of vapor present in the atmosphere.

It has been shown by Aitken that condensation sets in always around particles of foreign substances[1] (dust), floating in the air,

[1] We shall see later (in Radioactivity) that certain electrical effects, ionization by ultra-violet light, Roentgen rays or Becquerel rays, also furnish nuclei for condensation.

and that dust free water vapor may be considerably supercooled without condensation taking place.

The size of the drops formed depends upon the number of nuclei present. The more numerous they are the smaller the drops. The rate with which the latter sink through the air depends greatly upon their size. The drops in a fog are smaller than raindrops and the dense fogs in large cities are explained by the large number of dust and soot particles in the air.

The condensation of water vapor can easily be shown. Place a dish containing water under the bell-jar of an air pump and let the air get saturated with the vapor. A stroke of the pump will suddenly decrease the pressure and the vapor will be cooled below the dew point (see § 304); a fine mist of condensed moisture appears in the jar. Beautiful color effects are produced if a bright light is placed behind the jar.

Formation of clouds necessarily results if vapor laden air moves into regions of lower temperature and less pressure. If the temperature falls below the freezing point before the dew point is reached ice particles are formed instead of drops. Halos are optical phenomena produced by such ice crystals floating in the air, and the corona is a similar effect due to water drops of small size.

297. Hygrometry. Hygrometry is the measurement of the amount of water vapor in the atmosphere. The latter always contains some water vapor, but as a rule, is not saturated with it, so that evaporation takes place more or less freely from all bodies of water, the rate depending upon the temperature, the dryness of the air and wind effects (§ 284).

298. Relative Humidity, or fraction of saturation, is the ratio of the pressure of the water vapor in the air to the pressure which saturated vapor would have at the same temperature, *i. e.*, to the vapor tension at the temperature of the air. The larger this ratio is the slower the evaporation; if it becomes unity no further evaporation is noticed. Our sensation of dryness or dampness depends upon the relative humidity and not upon the actual amount of vapor present.

Since the vapor tension increases rapidly with the temperature (§ 283), an amount of moisture which produces only a small relative humidity during a warm summer day may saturate the air

at a lower temperature. Thus a cubic meter of air under normal pressure can contain as a maximum only 4.8 grams of moisture at 0°, but 17 grams at 20° and 30 grams at 30°, or, expressed in different units, one kilogram of air saturated with moisture contains at 0° 3.75 grams, at 20° 14.33 grams and at 30° 26.18 grams of water.

299. Relative humidity is determined by instruments called hygrometers.

The *dew point hygrometer* (see Fig. 154) consists of a vessel containing a volatile liquid which can be rapidly evaporated and thus cooled. The temperature is measured by a thermometer placed inside the liquid.

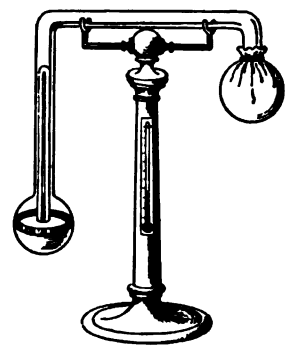


FIG. 154.

The metallic outer surface of the vessel is highly polished. As soon as the temperature has been lowered to the dew point, a mist is formed on this surface. The vapor tension corresponding to the dew point gives the actual vapor pressure of the moisture in the air; this divided by the vapor tension corresponding to the temperature of the air is the relative humidity. Both pressures are read from the tables (p. 228).

Other hygrometers are based upon the principle that certain substances, called hygroscopic, absorb moisture from the air and change their dimensions by this process. Moist substances consisting of organic tissue are hygroscopic; ropes and catgut strings shorten when moist, hair from which the oil has been removed, lengthens. A well known instrument of this type is Saussure's *hair hygrometer* (see Fig. 155).

Another principle is used in the August "*psychrometer*" or "wet and dry bulb thermometers" (Fig. 156). They consist of two similar thermometers the bulb of one of which is kept moist by means of a cotton or muslin wick surrounding it and dipping into a vessel of water. The constant evaporation on the surface of this thermometer lowers its temperature and the difference of

the readings of the two thermometers is a definite function of the relative humidity and temperature of the air.

Frequently the wet thermometer whose bulb is surrounded by muslin is simply moistened and then whirled through the air (sling thermometer) or a rapid stream of air drawn past it by aspiration (Assmann's aspiration psychrometer).

Different forms of psychrometers require different tables from which the relative humidity may be read off directly, if the tem-

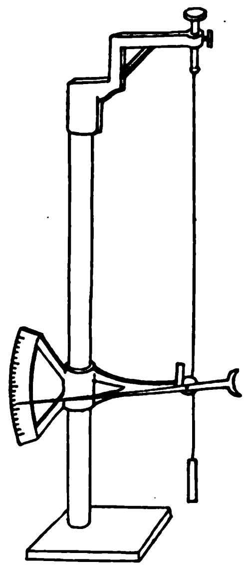


FIG. 155.

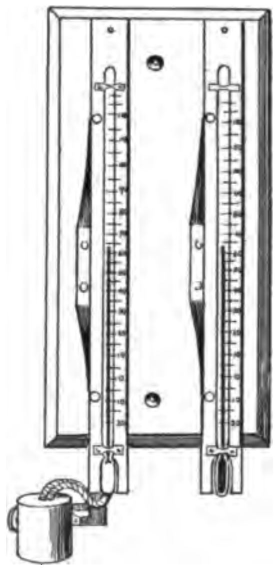


FIG. 156.

perature of the air and the difference in the readings of the two thermometers are given. The formulæ used for the calculation of these tables are quite complicated.

The *Absolute Humidity* is the amount of moisture contained in unit volume of air and can be determined by chemical means. A measured volume of air is passed over a substance absorbing water (calcium chloride or phosphor pentoxide). The increase in weight of this substance directly gives the amount of water absorbed.

LIQUEFACTION OF GASES.

300. Early Experiments. In the article on evaporation (§ 282) we have seen that in many cases a gas can be liquefied by either a decrease of temperature or by an increase in pressure.

Thus von Marum liquefied ammonia in 1799 by increasing the pressure to six atmospheres and Northmore in 1805 condensed hydrochloric acid, sulphur dioxide and possibly chlorine.

In 1823 Faraday began his well-known experiments upon the condensation of gases. He used a bent closed tube, containing in one arm a substance which upon being heated generates the gas to be liquefied, for example, sodium carbonate for the production of carbon dioxide. The other arm of the tube was placed in a freezing mixture. The gas developed increased the pressure in the tube sufficiently to liquefy it on the cold side.

301. The Critical Point. In 1863 Andrews enclosed liquid carbon dioxide in a glass tube so as to fill about one half of it. The tube was then carefully heated. The volume of the liquid increased though the vapor pressure rose considerably, the surface of demarcation between the two states became fainter, lost its curvature and finally disappeared at a temperature of 31°; the space was then occupied by a homogeneous fluid.

If the temperature was lowered again to 31° a mist appeared in the center of the tube and soon gave way to a well-defined meniscus showing that the homogeneous mass had separated again into liquid and gas. This experiment shows that the distinction between liquid and gaseous carbon dioxide ceases to exist at this temperature.

The effect produced by compressing carbon dioxide gas at different temperatures can be graphically shown by plotting the pressure as a function of the volume occupied by unit mass of the substance (Fig. 157). Let us follow the change taking place when the gas (point P) is compressed, the temperature always being kept constant (in the given example 20°). The volume decreases to a certain point, A, where condensation begins, and from that point on the pressure remains constant until the whole mass has been liquefied, at the point B; then further compression is accompanied by only a small decrease of volume. From A to B the two states are in equilibrium with each other.

Similar constant temperature lines, called isothermals, can be traced out for other temperatures.

The region in which a mixture of liquid and gas can exist and which is marked in the diagram by the dotted line becomes narrower as the temperature rises and ends at the point where the isothermal shows only a point of inflection. Above this no liquefaction, *i. e.*, the formation of two distinct states of different

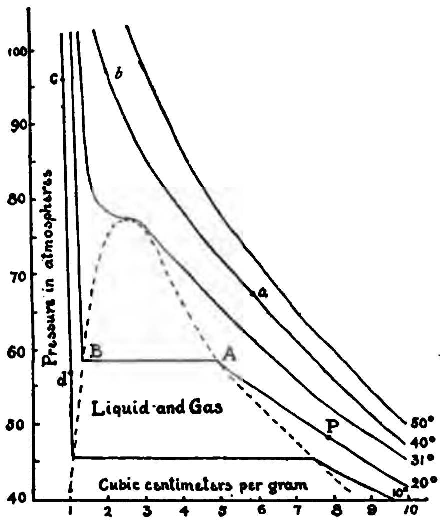


FIG. 157.

density, one a liquid, the other a gas, takes place. For still higher temperatures the curves approach more and more those of an ideal gas. The temperature to which a gas must be cooled before liquefaction by pressure is possible is called the *critical temperature.* The corresponding pressure and specific volume are called critical pressure and critical volume.

Fig. 157 is the projection upon the pressure-volume plane of a surface similar to the one represented in Fig. 153. The point at which the isothermal representing the critical temperature touches the liquid-vapor surface is denoted in Fig. 153 by H, and is called the *critical point.* It is, therefore, *the point of maximum pressure and temperature on the liquid-vapor surface.* In Fig 151 the point C, and in Fig. 152 the point at which the evaporation curve ends, represent the projections of the critical point

upon the volume-temperature and the pressure-temperature planes respectively.

It is idle to ask if the homogeneous fluid obtained in Andrews' experiment is a liquid or a gas. If a gas is first compressed at a temperature higher than the critical, then the temperature lowered and afterwards the pressure, keeping the latter higher than the vapor tension (*abcd* in Fig. 157), the substance has passed by a continuous process from a region distinctly belonging to the gas state to one in which it is distinctly a liquid. The continuity of the enclosed mass has never been broken. We can distinguish the two states only when the temperature is below the critical.

It has been proposed to restrict the term vapor to that part of the gas which lies below the isothermal through the critical point.

302. Continuity of State. Two years after the publication of Andrews' experiments James Thomson suggested that the two branches of each isothermal on either side of the region of mixture are parts of the same continuous curve which has a form as drawn for the isothermal of 20° in Fig. 158.

This would represent the path of the substance if it could pass continuously from one state to the other. The portions of the curve between

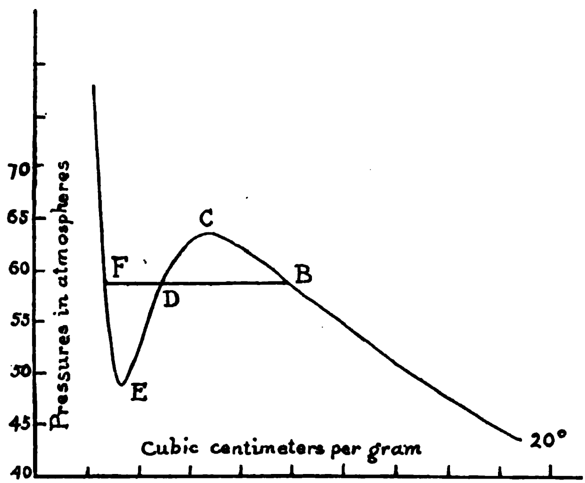


FIG. 158.

B and *F* have been realized experimentally for a short distance on either side. The possible existence of the part *CE*, where the volume would decrease with the decrease of pressure, can harldy be conceived. The straight line *BF* must be drawn, as Maxwell has shown, so that the areas bounded by the hypothetical curve above and below it are equal. For the

work necessary to bring the substance from point B to the point F must be the same along either isothermal (see § 56).

An attempt has been made by several physicists to derive equations of state for the liquid as well as the gaseous state. The best known of the large number of equations of this kind is the one, proposed by van der Waals in 1879

$$\left(P+\frac{a}{V^2}\right)(V-b)=RT \tag{73}$$

in which a and b are constants for a given substance. This equation agrees fairly well with Andrews' experimental results. The cubic equation for v has either one or three real roots. The one real root applies to that part of the diagram where only one volume corresponds to given values of p and T; the three roots to that region in which for a given temperature and pressure the substance is either a saturated vapor (B) or wholly liquid (F) or a mixture of the two. The real root representing the latter corresponds to the point where Thomson's hypothetical curve cuts the line of constant pressure in the region of mixture (D).

303. Critical Constants. While we have restricted ourselves in the preceding paragraphs to a discussion of the results obtained with carbon dioxide, the same phenomena occur with other liquids and gases. The former failures to liquify such gases as oxygen, nitrogen and hydrogen by the application of extreme pressures were due to the fact that in these experiments the temperature was not lowered below the critical value.

Usually the relations existing between volume, pressure and temperatures are expressed in C. G. S. units and degrees Centigrade. These units are arbitrary and possibly simpler relations may be found by selecting more "natural" units.

Van der Waals believes that the critical state is the same natural state for all substances and should be chosen as the reference point. In this case P_c, V_c and T_c corresponding to the critical point must be the units in terms of which pressure, volume and temperature should be measured.

CRITICAL CONSTANTS.

Substance	Critical Temperature.	Critical Pressure.	Critical Volume.
Water	365°	200 atoms	2.3 cc./g.
Alcohol	240	63	4.0
Ether	195	36	3.8
Carbon dioxide	31	77	2.2
Oxygen	— 118	51	1.5
Nitrogen	— 146	34	2.5
Hydrogen	— 242	20	—

In this new coördinate system each point of the diagram would represent "corresponding states" for all substances. In fact, Amagat has shown that the isothermals of carbon dioxide, air, ether and methane very nearly coincide if plotted in this manner.

304. Effect of Sudden Expansion. By cooling a gas below its critical temperature and applying a sufficient pressure it can be liquefied (§ 303). For several years after this was clearly recognized, oxygen, hydrogen, nitrogen and several other gases still resisted condensation even when cooled to the lowest temperatures obtainable at that time (solid carbon dioxide); they were therefore called permanent gases.

A new principle, allowing a liquefaction of these gases was found in the lowering of temperature which accompanies the sudden expansion of a compressed gas.

If a gas, contained in a vessel under a given pressure flows through a small opening into a space where the pressure is less it will produce a wind, or, in other words, it will impart to the particles passing the opening a considerably larger kinetic energy than they had before. This energy has to be supplied by the gas left in the vessel and consequently the temperature of this *remaining* gas must decrease; the kinetic energy is transformed into work, namely the work which the gas would have done by its pressure upon a piston in a cylinder (§ 230). If the expansion is so sudden that no heat is communicated from the walls of the vessel to the gas (adiabatic expansion) the fall in the temperature of the gas will be quite large.

Now let us suppose that the opening is made fine enough so that the passage of the gas through it becomes so slow that the average kinetic energy of the outflowing gas particles does not exceed that which they had before, and let the high pressure P_1 be kept constant by the continuous working of a compressor. After a state of steady flow has been established the work done upon the gas by the piston in forcing out a quantity, originally at a volume V_1, is P_1V_1. The work done by this quantity in expanding to a volume V_2 against a pressure P_2 is P_2V_2. If it receives no heat from the outside the change in the internal energy of the gas is represented by $(P_2V_2 - P_1V_1)$. If now in a given case the temperature of the gas does not change during the expan-

sion and if the gas is perfect $P_2V_2 = P_1V_1$, and there is no change in the internal energy. The gas itself does no external work but serves, so to speak, merely as a buffer between the piston of the compressor and the outer air. But if the temperature—this is the case for all gases except hydrogen—falls during the expansion we might reasonably expect that P_2V_2 should exceed P_1V_1 and while Boyle's law cannot be applied directly, it is a fact that for most gases the product PV decreases with increase of pressure (§ 222). Energy must therefore be added to these gases when they expand under the stated conditions; as it is usually explained, internal work must be done to overcome the attraction of the molecules for each other. This shows itself in a fall of the temperature of the gas *while it passes* through the opening. On the other hand, hydrogen gas for which under ordinary pressures and temperatures the product PV increases with the pressure, experiences a heating instead of a cooling. This was proved experimentally by Joule and Thomson (Lord Kelvin). They separated the two vessels by a plug of cotton wool, so that the gas would pass slowly from one to the other, and measured the temperature of the gas just before and after the passage. (Porous plug experiment.)

In general both effects mentioned above must be considered when a gas is made to expand suddenly, and both, except in the case of hydrogen, will produce a cooling of the gas.

Applying this result to a highly compressed gas, contained in a cylinder and cooled to a low temperature which, however, is still higher than the critical, the latter may be reached by an additional sudden expansion of the compressed gas.

Oxygen was first liquefied by Cailletet in 1877, and in the same year by Pictet, by suddenly releasing a pressure of 300 atmospheres on the gas cooled to — 29 degrees. Other permanent gases were later liquefied by the same method, but it is clear that the yield is only small and the process cannot well be made continuous.

305. The Regenerative Process. Linde, and at about the same time, Hampson, discovered in 1895 a method for the continuous liquefaction of the so-called permanent gases. By means of a compressor the gas is kept under 150 or 200 atmospheres, cooled

to a low temperature and then allowed to stream out through a small opening into a vessel in which the pressure is about 16 atmospheres or less. The gas cooled by expansion is led back through spiral copper tubes which are in intimate thermal contact with the compressed gas flowing towards the orifice. The temperature of the gas before it reaches the opening is thus reduced more and more by the cooler counter current, until finally it sinks to the point at which liquefaction begins under the lower pressure in the vessel into which the gas streams. After this the liquid is produced continuously. Thus liquid air can be manufactured in large quantities and in its turn may be used for the initial cooling of gases that are still more difficult to liquefy.

Fig. 159 represents Linde's apparatus for the liquefaction of air. A compressor forces the air under a pressure of 200 atmospheres into the tube p_1, through the cooler g and at p_1' into the innermost of a system of three concentric spiral tubes. The gas flows down through this tube to the expansion valve a where the pressure is reduced to 16 atmospheres. The air cooled by expansion passes up the middle tube and finally back through p_2 to the compressor. Below a there is another expansion valve b reducing the pressure to a little above one atmosphere, lowering the temperature still further. The cold air below b passes through the outer tube and finally escapes through p_3.

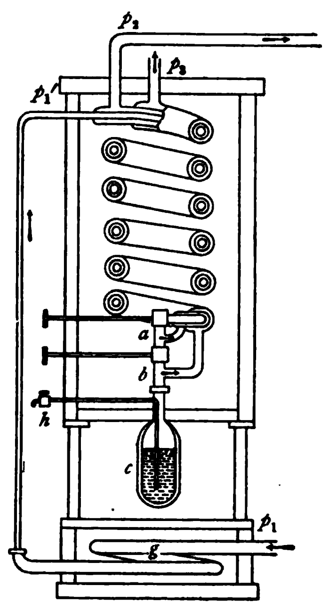


FIG. 159.

An auxiliary compressor connected with the pipe p_2 supplies some air from the outside to replace that lost through p_3 and keeps at the same time the pressure in p_2 at 16 atmospheres.

After the apparatus has been working for a little over an hour the counter currents of cold air in p_2 and p_3 have reduced the

temperature of the air in the inner tube sufficiently so that upon expansion at the valve *a* some will liquify and pass into the "Dewar flask" *c*. This is a double-walled glass vessel, the space between the walls being highly evacuated; thus heat is prevented as much as possible from reaching the interior of the vessel. The liquid air can be drawn from the collecting flask by means of the stopcock *h*.

Liquids boiling at very low temperatures must evidently be carefully insulated thermally; they can be kept in Dewar flasks for several days.

Hydrogen, which at ordinary temperatures forms an exception to the general rule (last section), cools upon expansion, if the temperature is low enough (—80° according to Olzewsky). Dewar succeeded in 1898 in liquefying hydrogen by the regenerative process.

The gas was cooled by passing it through baths of liquid carbon dioxide, liquid air and, finally, of liquid air boiling under reduced pressure (—200°); then it was allowed to expand from a pressure of 150 atmospheres to one atmosphere.

By boiling liquid hydrogen under reduced pressure Dewar froze it to a foam-like solid at a temperature of —258°.

MELTING POINTS, BOILING POINTS AND CRITICAL TEMPERATURES.

Substance.	Melting Point.	Boiling Point.	Critical Temperature.
Helium	——	—268.5°	—266°
Hydrogen	—258°	—252	—242
Nitrogen	—210	—195.5	—146
Air	——	—191	—140
Oxygen	—227	—183	—118
Fluorine	—223	—187	—120
Carbon monoxide	—207	—190	—141
Carbon dioxide	— 65	——	+ 31
Hydrochloric acid	—116	— 35	+ 51.3
Ammonia	— 77	— 33.7	+130
Chlorine	—102	— 36.6	+141
Sulphur dioxide	— 76	— 8.0	+155.4

Kamerlingh Onnes in Leiden succeeded (1908) in liquefying helium. The gas, under a pressure of 100 atmospheres, was cooled in a bath of boiling hydrogen to —258° and its temperature still further lowered by the regenerative process. The boiling point

of helium under atmospheric pressure is about $-268^\circ.5$. By boiling it under reduced pressure, corresponding to 1 cm. of mercury the temperature is lowered to -270° or 3 degrees above absolute zero.

RADIATION.

306. Nature of Radiation. In § 269 it was mentioned that an exchange of heat may occur between two bodies without a heating of the intervening medium and that this is called "radiation."

Unless heat is produced by some independent process the body radiating into space is losing heat energy and cools. The energy thus sent out is no longer heat and cannot be measured directly by any of the methods studied in the chapter on calorimetry. The temperature of bodies which allow it to pass through them and which are called transparent to it is not affected; but those substances which absorb it, *i. e.*, are opaque to it, are warmed because the radiant energy is retransformed into heat. The heating effect of the sun is noticed at the same moment at which the light reaches the earth. The radiant energy passes therefore through a vacuum with the same velocity as light: it travels in straight lines, casts shadows (fire screens); its direction is changed by mirrors, prisms or lenses; in short, in all its physical properties it is indistinguishable from light, and the meaning of "radiation" must be enlarged so as to include them both and possibly some other similar phenomena (see Light).

Radiation is explained as a vibratory disturbance or wave motion of an imponderable substance which is assumed to fill all space and is called the ether.

In a wider sense radiation may be taken as propagation of energy of any form through space, but we shall in the following restrict ourselves to ether radiation and study it only as far as it is related to heat phenomena.

307. Radiometers. A radiometer is an instrument which is used to detect and measure, by the heat effects produced, the amount of radiant energy falling upon it. The instrument should absorb all radiation reaching it and neither transmit nor reflect any. Lampblack answers this purpose well and for this reason surfaces of radiometers are covered with it (see § 309).

A thermometer with a blackened bulb may be used, but it is not very sensitive. A very effective radiometer is the "*Thermopile*" which consists of a large number of thermoelements (§ 248*c*) connected as shown in Fig. 160.

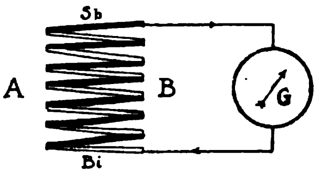


FIG. 160.

If all the junctions on one side, for example *B*, are kept at the same temperature and the temperature on the other side, *A*, differs from it, the current through a galvanometer connected with the terminals of the pile will flow in one direction or the other, according as *A* is warmer or cooler than *B*. The whole instrument is therefore nothing but an extremely sensitive thermometer, indicating the temperature of *A*, the position of rest of the galvanometer corresponding to the temperature at which *B* is kept. A complete thermopile consisting of a number of rods of bismuth and antimony, which metals have a large thermoelectric effect, is shown in Fig. 161.

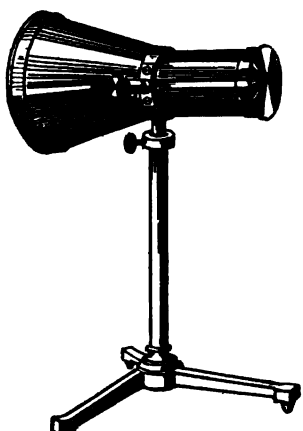


FIG. 161.

The bolometer, brought out in 1881 by Langley, makes use of the change of electrical resistance due to heating. A fine grating of thin metallic strips or single strip of platinum, blackened on the surface, is exposed to radiation. This metallic grating or strip is placed in one arm of a Wheatstone bridge (see Electricity) and the deflections of the galvanometer in the bridge become a measure of the rise in temperature due to the absorbed radiations and consequent rise in resistance of the strip.

308. Theory of Exchanges. Hold a body warmer than the room in front of one face of a thermopile. This is then heated by radiation and the galvanometer shows a deflection. A burning match will affect the instrument at quite a distance. If a piece of ice is substituted for the match, the deflection is in the opposite direction showing that the side of the thermopile which before absorbed energy now radiates and is cooled.

We can hardly assume that in the first case the surface of the thermopile sent out no radiation at all, but it is more probable that it absorbed more than it lost and therefore was heated, while in the second case it was cooled, because it received less than it sent out.

From this point of view there is a continuous interchange of energy of radiation between all bodies and the heating or cooling depends only upon the differential effect of the radiation absorbed and emitted; if the two are equal the temperature remains constant.

This theory was first enunciated by Prevost in 1792 and has been of the greatest value for a correct interpretation of radiation phenomena.

If bodies originally at different temperatures are enclosed in a vessel which neither receives heat from nor loses heat to the outside, an equilibrium of heat exchanges must result finally, *i. e.*, all bodies in the enclosure as well as the walls of the vessel will be at the same temperature.

309. Emission and Absorption of Radiant Energy. Leslie showed that the emission of radiant energy from bodies at the same temperature differs considerably with the nature of the surface. He used a cube of block tin (Leslie's cube) filled with hot water and having its surfaces coated with different materials. Turning one side after another towards the thermopile he proved that polished metals radiate much less than a surface covered with varnish, and this again less than one covered with lampblack. In general bad radiators are good reflectors, and, if a body reflects a large amount of the radiation incident upon it, it absorbs little.

A tabulation of various substances according to their radiating and absorbing power shows that both of these vary in the same sense.

The following experiment may illustrate this. Draw a figure with India ink upon a sheet of platinum and hold it in a Bunsen flame, with the ink mark towards the flame. The figure will appear dark on a bright ground, since it radiates more heat than the rest of the sheet and the platinum in contact with it is therefore cooler. But if the sheet is reversed the platinum is heated by the flame equally throughout, and the mark appears much brighter than the rest on account of its larger radiating power.

The absorptive power of a surface is the ratio of the quantity of radiant energy absorbed to the total quantity falling upon it. A body which absorbs all radiation, *i. e.*, whose absorptive power is unity, is called a "perfectly black" body or simply a "black" body, an expression introduced by Kirchhoff. In such a body all radiant energy is transformed into heat. A small hole in a hollow closed vessel is a near approach to such a black body. Practically all radiation passing into the hole is reflected internally and finally absorbed.

Taking the black body then as a standard, the emissive power of any surface is defined as the ratio of the energy emitted by it to that emitted by a black body at the same temperature.

310. Light is Radiant Energy. The Spectrum. As mentioned in § 306 there is no difference between light and radiant energy. If we pass a beam of white light through a prism, a band of colored light, called a spectrum, is produced, ranging from red through orange, yellow, green, and blue to violet. We shall see later (see Light) that this "breaking up" of the white light is due to the fact that different colored lights have a different wave length and that the wave length of the extreme red in the spectrum is 0.76μ or 0.00076 mm. and that of the extreme violet 0.4μ, the wave length constantly increasing from the violet to the red.

With a thermopile placed at the violet end of the spectrum an effect can be observed only with the most sensitive instruments, showing that the energy of radiation is very small. But as we approach the red end the effect upon the thermopile becomes much greater. However it does not stop at the extreme red but is noticeable far beyond it, and from this the conclusion must be drawn that ether waves still longer than red light exist, though not visible to the eye. Light is then the visible part of radiant energy. While the eye can distinguish the quality of radiation as color, a radiometer measures only the intensity or energy, not the quality.

Formerly a distinction was made between "light rays" and "heat rays," but apparently the distinction exists only in the limitations of our eye, not in the nature of the rays.

311. Selective Transmission and Absorption of Radiant Energy. Red glass transmits only such rays of the visible spectrum as give

the effect of red light while the rest is absorbed. This selective absorption of rays of certain wave length is a common phenomenon in nature. The effect is in general a production of heat.

The relation between radiating and absorptive power demands that the bodies which absorb rays of certain wave length will radiate these rays more strongly than others. A piece of red glass which absorbs the green and blue rays, will shine in the dark, after being heated in a Bunsen flame, with a bluish light.

The difference between absorption and radiation is then simply a question of relative intensity of vibration. If that of the incident rays is the greater, absorption takes place, but if the molecular vibrational energy of the absorbing body is sufficiently increased, for example, by heating, it will in general send out the very rays which it absorbed at a lower temperature.

The wave length of radiation is therefore of prime importance in the discussion of the transparency of a medium. Investigations reveal the fact that bodies which are transparent to visible rays, often absorb the longer waves. Such a substance is glass. The large deflection produced by a Bunsen burner placed at some distance from a thermopile is considerably decreased if a piece of glass is placed in front of the instrument. The protective quality of window glass against loss of heat from the rooms is due to this property. It admits light and its energy unhindered into the buildings, but does not permit the longer rays to pass out.

Glass is unsuited for instruments used to investigate the properties of rays of long wave length and must be replaced by rock-salt or fluorite, which are transparent to them.

Other substances, as thin sheets of vulcanite or a solution of iodine in carbon bisulphide are opaque to light but transparent to longer waves.

Bodies which transmit radiation of great wave length are called diathermanous, those which absorb them, athermanous, though these terms are necessarily very indefinite without special reference to wave length.

Liquids and gases also show a considerable difference in their absorptive power for long waves. Thus while carbon bisulphide transmits them, water absorbs them strongly. Solutions fre-

quently absorb less than the solvent, for example, alum solution transmits more than pure water.

Air is highly diathermanous but water vapor and carbon dioxide absorb, and it is due to the presence of these gases in the atmosphere that a large part of the sun's energy is stopped before it reaches the surface of the earth.

312. Reflection and Refraction of Radiant Energy. Investigation has shown conclusively that the long invisible waves are reflected and refracted in the same manner as the shorter light waves. A more complete discussion of these phenomena is left until later (see Light).

(*a*) Parallel rays falling upon a concave spherical mirror converge after reflection towards a point, called the principal focus and situated at a distance of half the radius of the sphere from the center of the mirror. On the other hand, if a source of radiant energy is placed at the principal focus of such a mirror the rays after reflection are parallel to the axis of the mirror.

If a heated iron ball is placed in the focus of one mirror (Fig. 162) and a thermopile in the focus of another, whose axis is coincident with that of the first, the waves from the ball will after travelling to the second mirror converge on the thermopile and produce strong heat effects which disappear as soon as the thermopile is moved away from the focus. The surface of the mirrors

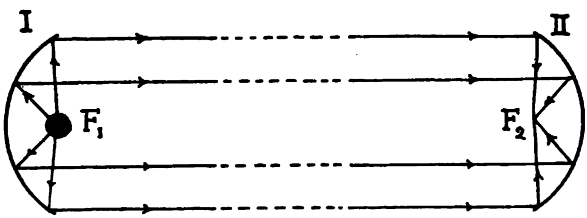


FIG. 162.

does not need to be highly polished as would be necessary for a corresponding experiment with light waves.

(*b*) If a thin flask filled with a solution of iodine in carbon bisulphide is placed in the rays from an arc light it will act upon them as an ordinary lens would, *i. e.,* converge them to a focus a short distance behind the glass. Though no light is transmitted

through the flask, the longer waves pass, and if a match is held at the focus it can easily be lighted by the heating effects of these rays.

313. Influence of Temperature upon Radiation. It is a common experience that the intensity of radiation increases with the temperature. Newton's law of cooling states that "The rate of cooling of a body is proportional to the difference between its own temperature and that of the surroundings," but this law holds only for small temperature differences and includes cooling both by radiation and convection. Stefan discovered in 1879 that the rate of radiation from a body is proportional to the difference of the fourth powers of the absolute temperatures of the body and of the surrounding medium.

If radiation takes place between two bodies at the absolute temperatures T_1 and T_2, where $T_1 > T_2$, Stefan's law takes the form

$$R = c(T_1^4 - T_2^4), \tag{74}$$

where c is a proportionality factor.

The above equation may be written:

$$R = c(T_1 - T_2)(T_1^3 + T_1^2T_2 + T_1T_2^2 + T_2^3) \tag{75}$$

or if the temperature difference is small so that T_2 nearly equals T_1,

$$R = 4cT_1^3(T_1 - T_2) = C(T_1 - T_2) \tag{76}$$

which is Newton's law of cooling.

Each square centimeter of a body covered with lampblack, kept at 100° and surrounded by a shell kept at 0°, loses one calorie per minute, if the intervening space is a vacuum.

Stefan's law holds strictly only for perfectly black bodies which according to definition (§ 309) are free from selective absorption or radiation. The radiation from a black body is independent of any substance and a function of the temperature only; for example, that from a small hole in an enclosure with walls at a uniform temperature does not depend upon the nature of the walls, or any bodies within the enclosure.

This conclusion was reached independently by Kirchhoff and Stewart about 1858.

314. Distribution of Energy in the Spectrum. When radiation proceeds from a black body, rays of all wave lengths are present and a continuous spectrum can be formed in which the energy of each portion of the spectrum is a definite fraction of the whole. Such radiation is called "normal," or "black body" radiation.

Plotting the rate of radiation per square centimeter (J) as a function of the wave length, the curves in Fig. 163 are obtained;

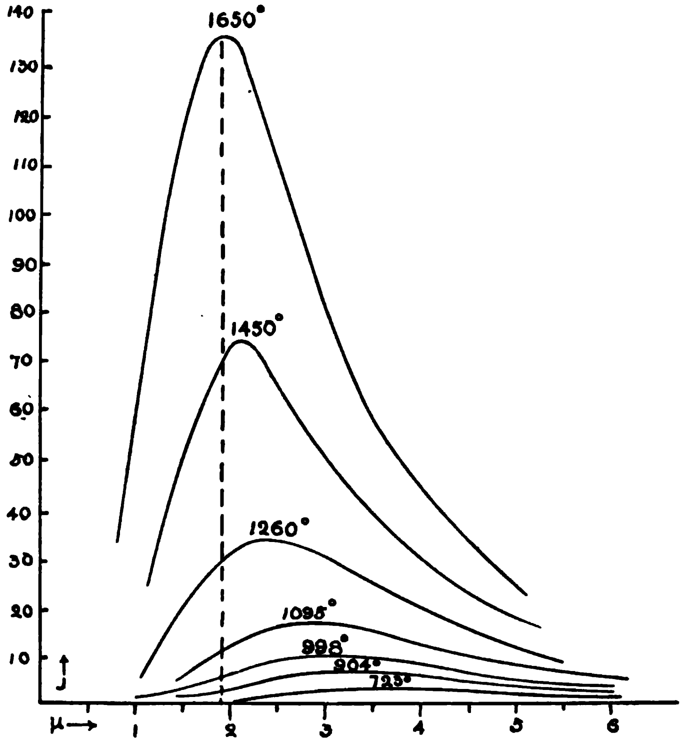


FIG. 163.

in each curve the maximum corresponds to a definite wave length. The unit of length employed in Fig. 163 is 1 μ or 0.001 mm. As the temperature is raised this maximum shifts towards shorter wave lengths and the radiation contains more rays from the bluish part of the spectrum or the color becomes whiter.

Wien has shown that for a body emitting normal radiation

$$\lambda_m T = \text{constant} \qquad (77)$$

where λ_m is the wave length corresponding to the maximum of the energy curve. This law is called Wien's displacement law; in general the constant varies with the surface and temperature of different radiating bodies. This subject will be discussed more fully under "Light."

315. Measurement of Temperature by Radiation. After the constants in equations (74) or (77) have once been determined the equations can be used for the measurement of the temperature of radiating bodies, either by measuring the rate of radiation or by locating the wave length corresponding to the energy maximum in the spectrum.

The measurement of temperatures by this method is called optical pyrometry.

Thus it is possible to measure approximately the temperature of bodies beyond our reach or so hot that other methods are inapplicable. It should, however, be kept in mind that the above equations hold strictly only for black body radiation; the "black body" temperature calculated from the formula may be somewhat smaller than the real temperature. The differences are, however, relatively not very large for most bodies, but may amount to several per cent.

In this manner the "black body" temperature[1] of the sun has been found to be 6000° and that of the electric arc to be about 3500°. The crater of the arc is the hottest part and the temperature increases about 100° if the current is increased from 15 to 30 amperes. The temperature of the Nernst lamp is 1950°, of an incandescent lamp 1500° and of the Bunsen burner about 1750°.

THERMODYNAMICS.

316. The Mechanical Equivalent of Heat. We have already become familiar with the fact that heat is a form of energy and that mechanical energy can be transformed into heat and *vice versa,*

[1] The real temperature of the sun's surface is estimated by different investigators to be between 7000° and 10000°.

but we have usually employed a different unit for quantity of heat from that for mechanical energy.

Just as it was shown in Mechanics that in purely mechanical processes the total energy remains constant, so in the section on calorimetry we have learned that the total quantity of heat remains constant in a system of bodies between which purely thermal processes take place. (Law of equal heat exchanges, § 263.)

A great step forward was made when it was found that whenever mechanical energy is transformed into heat there exists always a definite relation between the *numerical* expressions for these two forms of energy. This may be written as a mathematical equation thus:

$$W = JH, \tag{78}$$

where W is expressed in mechanical units and H in heat units. The constant ratio J between W and H is called *the mechanical equivalent of heat*. It may be defined as the number of units of mechanical energy that are required to produce a unit of heat, *e. g.*, the number of ergs that will produce a calorie.

Joule's classical determinations of this important constant consisted in measuring the heat developed and the mechanical work spent, when water was stirred by means of paddle wheels.

In 1878 and 1879 Rowland repeated Joule's experiments with greatly improved apparatus. The paddle wheel was driven by a motor. The water was contained in a calorimeter with stationary vanes to prevent a rotation of the whole mass of water. The calorimeter was not fixed but suspended by a fine wire, so that it could rotate around a vertical axis. It was fastened to a circular disk of radius R, concentric with the axis of the calorimeter.

To counteract the mechanical moment produced by the rotating paddles two strings were attached to diametrically opposite points of the disk and tangential to it. They then passed over pulleys and supported weights sufficiently large to keep the calorimeter from turning.

The moment due to these weights, each of mass m, is then

$$\text{Moment} = 2RF = 2Rmg \text{ dyne cms.} \tag{79}$$

The work against this couple during one revolution of the paddle wheel is (§ 93)

$$w = 2\pi \cdot 2Rmg \text{ ergs} \tag{80}$$

and during n turns,

$$W = 4\pi nRmg \text{ ergs} \tag{81}$$

This work is transformed into heat. If the water equivalent of the calorimeter with contents is C and the rise in temperature t

$$H = Ct \text{ calories.} \tag{82}$$

But if we wish to express the heat in ergs instead of in calories, we must write

$$W = JCt \text{ ergs} \tag{83}$$

and

$$J = \frac{4\pi nRmg}{Ct} \, \frac{\text{erg}}{\text{calorie}} \tag{84}$$

Rowland's result for J reduced to the hydrogen scale gave

$$\text{One calorie} = 4.187 \times 10^7 \text{ ergs. at } 15°. \tag{85}$$

Since that time the mechanical equivalent has been repeatedly determined, also by electrical methods (see § 318). All results agree closely with Rowland's; in fact the above value is the average of the best recent determinations.

It is apparent that the value of the mechanical equivalent of heat depends upon the units chosen. Thus if we select the kilogram-meter instead of the erg, and the large calorie, K, as heat unit, J will be 427 (kilogram-meter)/(large calorie). Using the English system we obtain for one British thermal unit

$$\text{One B. T. U.} = 778 \text{ foot pounds.} \tag{86}$$

While in all preceding work the calorie was used as the unit of heat, because it is most convenient for heat measurements, quantity of heat can as well be expressed in mechanical units as in thermal units.

317. The First Law of Thermodynamics. The fact that quantity of heat and mechanical work are equivalent, can be formulated

thus: "When equal quantities of mechanical effect are produced by any means whatever from purely thermal sources or are lost in purely thermal effects, equal quantities of heat are put out of existence or are generated" (Kelvin). The steam engine is the best illustration of the production of mechanical effects from thermal sources, while mechanical agencies give rise to thermal effects when a body is heated by friction or impact.

This principle is called the first law of thermodynamics. Its mathematical expression is:

$$\Sigma W = \Sigma H \tag{87}$$

where ΣH means heat transformed into work and ΣW the total amount of work performed or mechanical energy produced at the expense of the heat so transformed, both being expressed in the same units. Hence the total change of energy, for a given transformation of heat into work or *vice versa,* is zero.

The formula is sometimes stated differently. If we consider W to be positive when it represents *external work done by* a body or system and H positive when it means an *increase* of the heat in the body or system, the sign of ΣH in (87) must be reversed and we get

$$\Sigma W + \Sigma H = 0 \tag{88}$$

318. Law of Conservation of Energy. After the first law of thermodynamics was once established it was only a step to extend it to other branches of physics. Mayer gave in 1845 the first clear statement of the law of conservation of energy, which was followed two years later by the classical paper of von Helmholtz: "Ueber die Erhaltung der Kraft." This most important physical law can be worded in many different ways, for example:

"The total amount of energy is any system of bodies is not changed by any reaction or transformation between the parts of the system."

Extending this principle beyond the field of actual experimentation it has also been stated thus: "Energy can be neither created nor destroyed."

Since we have to deal only with the transition of a system of bodies from one state into another and are unable to determine

the absolute amount of energy in a body, a more practical statement of the law of conservation of energy would be "The energy of a system in a given state, referred to a fixed (normal) state, has a definite value, independent of the way in which the transformation from one state to the other has taken place."

For suppose that there were two ways possible in which the system could pass from state A to state B and that in one the energy change would be different from that in the other. We could then let the system pass from A to B in one way but return to A in the other. In the end the system would be exactly in the same condition as at the beginning, but in addition energy would either have been created or destroyed.

The law of constant heat sums in chemical reactions (§ 267) follows directly from this principle.

The electrical experiments for the determination of the mechanical equivalent of heat also furnish a good illustration of the law of conservation of energy. We shall see later that electrical energy is equivalent to mechanical energy and can be measured in ergs; further that heat is produced in a wire in which electrical energy is consumed. By measuring the heat produced during the disappearance of a known amount of electrical energy, the value for the mechanical equivalent of heat can easily be calculated. Such determinations of J have been made by Griffiths, Schuster and Gannon, and Callendar and Barnes and, as stated above, the results agree very closely with the values obtained by the mechanical methods.

319. The Two Specific Heats of Gases. If a gas is heated under constant volume the heat necessary to raise its temperature from t_1 to t_2 is (§ 260)

$$H_v = Mc_v(t_2 - t_1) \tag{89}$$

If the gas is heated under constant pressure through the same temperature interval, its volume increases and the heat absorbed is

$$H_p = Mc_p(t_2 - t_1) \tag{90}$$

where c_v and c_p are the heat capacities of the gas under constant volume and constant pressure respectively.

In the second case work is performed by the gas, namely

$$W = P(V_2 - V_1) \tag{91}$$

Hence a greater amount of heat must be absorbed in the second case than in the first. If it is assumed that the expansion of the gas alone, *i. e.*, the separation of its molecules, requires no work (see § 304) the first law of thermodynamics gives

$$Mc_p(t_2 - t_1) - Mc_v(t_2 - t_1) = P(V_2 - V_1) \tag{92}$$

or

$$c_p - c_v = \frac{P(V_2 - V_1)}{M(t_2 - t_1)} \tag{93}$$

If the gas is a perfect gas the gas law (§ 256) may be applied,

$$P(V_2 - V_1) = R_1(T_2 - T_1) = R_1(t_2 - t_1) \tag{94}$$

and

$$c_p - c_v = R_1/M = R_2 \text{ cal. per gram-degree} \tag{95}$$

Of course c_p, c_v and R_2 must be expressed in the same units. If the mechanical units are used, the last equation becomes

$$J(c_p - c_v) = R_2 = PV/MT \text{ ergs per gram degree.} \tag{96}$$

The difference of the specific heats, $s_p - s_v$, is numerically equal to the difference of the heat capacities, $c_p - c_v$.

The last equation may be used to calculate the mechanical equivalent of heat, which was done first by Mayer in 1842. Solving for J we obtain

$$J = \frac{PV}{M(c_p - c_v)T} \tag{97}$$

Under atmospheric pressure ($P = 1012630$ dynes/cm^2.) and at 0° ($T = 273$) the density of air is 0.001293 gram per cm^3; $c_p = 0.238$ and $c_v = 0.169$. Therefore,

$$J = \frac{1012630}{0.069 \times 0.001293 \times 273} = 4.16 \times 10^7 \frac{\text{erg}}{\text{calorie}}$$

In the derivation of the formula it was assumed that the internal work done during the expansion of a gas is zero. Joule and Thomson proved by the porous plug experiment that, while it is not zero, it is very small in the case of permanent gases.

320. Ratio of the Specific Heats of Gases. In the section on the

molecular theory of gases (§ 227) it was shown that the molecular kinetic energy of translation of a mass M of a gas is

$$\text{Kin. En.} = \tfrac{1}{2}M\overline{v^2} \tag{98}$$

where $\overline{v^2}$ is the mean square of the velocities of the molecules; and that

$$PV = \tfrac{1}{3}M\overline{v^2},$$

or

$$P = \tfrac{1}{3}\rho\overline{v^2} \tag{99}$$

where ρ is the density of the gas.
From this follows

$$\frac{P_1V_1}{P_2V_2} = \frac{\overline{v_1^2}}{\overline{v_2^2}} \tag{100}$$

By Charles law, combined with the above equations

$$\frac{P_1V_1}{P_2V_2} = \frac{T_1}{T_2} = \frac{\overline{v_1^2}}{\overline{v_2^2}} = \frac{\tfrac{1}{2}M\overline{v_1^2}}{\tfrac{1}{2}M\overline{v_2^2}} \tag{101}$$

The kinetic energy of a gas is therefore proportional to the absolute temperature.

The heating of a gas at *constant volume* may produce two effects, 1, an increase of the molecular motion of translation changing the mean square from $\overline{v_1^2}$ to $\overline{v_2^2}$; 2, an increase of the internal energy of the molecules, *i. e.*, rotational and possibly vibrational energy. The total heat necessary to heat a gas under constant volume consists thus as a rule of two parts,

$$H_v = H_1 + H_2 = \tfrac{1}{2}M(\overline{v_2^2} - \overline{v_1^2}) + H_2 = c_vM(t_2 - t_1) \tag{102}$$

But if the heating takes place under constant pressure, an additional amount of heat must be added to H_v to take care of the external work. But, $PV = \tfrac{1}{3}M\overline{v^2}$, and therefore the work done during expansion, *i. e.*, $P(V_2 - V_1)$, equals

$$\tfrac{1}{3}M(\overline{v_2^2} - \overline{v_1^2}) = \tfrac{2}{3}H_1 \tag{103}$$

Thus heating through the same temperature range under constant pressure requires

$$H_p = \tfrac{5}{3}H_1 + H_2 = c_pM(t_2 - t_1) \tag{104}$$

Dividing (104) by (102) we obtain for the ratio of the specific heats (or heat capacities)

$$\gamma = \frac{c_p}{c_v} = \frac{\tfrac{5}{3}H_1 + H_2}{H_1 + H_2} \tag{105}$$

$$= 1 + \tfrac{2}{3}\frac{H_1}{H_v} \tag{106}$$

where $H_v = H_1 + H_2$ (see (102)), and

$$\frac{H_1}{H_v} = \tfrac{3}{2}(\gamma - 1) \tag{107}$$

H_1/H_v is the ratio of the increase of the translational energy of the molecules to the total increase in energy.

In any case

$$0 < H_1/H_v < 1 \tag{108}$$

and hence

$$1 < \gamma < 1.66 \tag{109}$$

If H_1/H_v is unity or if γ equals 1.66, the increase of molecular energy is wholly energy of translation and there is either no internal energy present or it is independent of the temperature

The former may be assumed to be the case in monatomic gases, *i. e.*, gases whose molecules consist of a single atom only. Since γ has been found to be 1.66 for argon, helium and mercury vapor these gases are probably monatomic.

In other cases the internal energy is an appreciable fraction of the whole, and its increase, H_2 of the preceding equations, greatly influences the amount of heat necessary for the heating of the gas. The more complex the structure of the molecule the more apparent will be the effect of the internal energy, while γ at the same time approaches the lower limit, unity (see Table p. 212).

321. Isothermal Processes. Any process in which the temperature remains constant is called an isothermal process. The simplest case is the isothermal expansion or compression of a perfect gas. If we represent the state of such a gas on the pressure volume diagram the isothermal lines, characterized by the equation

$$PV = R_1T = \text{constant} \tag{107}$$

are rectangular hyperbolas (see Fig. 164).

If a gas of volume V_1 under the pressure P_1 expands isothermally to a volume V_2 under pressure P_2 it performs external work and must absorb an equivalent amount of heat from the outside to prevent being cooled. The work done by the gas is represented by the area included between the curve, the volume axis and the two ordinates P_1 and P_2, the area shaded in the figure (§ 56).

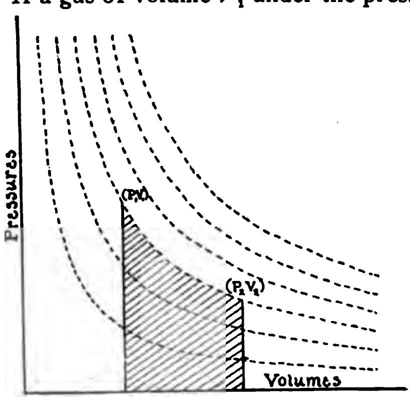


FIG. 164.

On the other hand if the gas is compressed isothermally between the same limits, the same amount of work is expended upon the gas and an equivalent amount of heat given out.

322. Adiabatic Process. A process in which the body (or system of bodies) under consideration neither receives heat from nor loses heat to other bodies is called an adiabatic process.

In the case of a perfect gas the equation for an adiabatic expansion or compression is (as shown in § 323)

$$PV^{\gamma} = \text{constant.} \tag{110}$$

Here also external work is done during expansion, and since no heat is supplied from the outside the gas must cool. The slope of an adiabatic line through a given point of the diagram is therefore steeper than that of an isothermal line, since in the former the pressure falls more rapidly.

An adiabatic compression results in a heating of the gas. The mechanical work done is again represented by the area between the curve, the volume axis and the two extreme ordinates.

Adiabatic processes are found in nature in rapid periodic vibrations, as those of sound in air, and for this reason they enter in the calculation of the velocity of sound (see "Sound").

323. Equation of an Adiabatic. Let an original volume V be expanded adiabatically to $V + dV$. The general expression for the heat required to change the temperature of a gas and let it perform external work, is from § 319

$$dH = Mc_v dT + \frac{PdV}{J} \text{ calories} \tag{111}$$

Since in the present case no heat has been supplied from the outside,

$$PdV + JMc_v dT = 0 \tag{112}$$

The gas law gives by differentiation

$$PdV + VdP = R_1 dT \tag{113}$$

Substituting the value for dT in (112)

$$PdV + \frac{JMc_v}{R_1}(PdV + VdP) = 0 \tag{114}$$

Since from equation (95)

$$\frac{R_1}{M} = J(c_p - c_v), \tag{115}$$

$$(c_p - c_v)PdV + c_v(PdV + VdP) = 0 \tag{116}$$

$$c_pPdV + c_vVdP = 0 \tag{117}$$

or writing

$$\gamma = \frac{c_p}{c_v}$$

$$\gamma \frac{dV}{V} + \frac{dP}{P} = 0 \tag{118}$$

Integrating

$$\gamma \log V + \log P = \text{constant} \tag{119}$$

and

$$PV^\gamma = \text{constant} \tag{120}$$

If we write (118) in the form

$$\frac{dP}{-\frac{dV}{V}} = \gamma P \tag{121}$$

we see that the adiabatic modulus of elasticity of gas, that is, its modulus of elasticity of volume (§ 169) when it is compressed adiabatically, is γP or γ times greater than its isothermal modulus (§ 223).

324. Carnot's Cycle. It is clear that an isothermal or adiabatic expansion, followed by a simple reversal will in the end leave no change whatever in the system. In order to obtain mechanical work from a gas and at the same time return it to its original condition, so that the process may be repeated at will, the expansion of the gas and its return to the starting point must proceed along different paths.

An ideal cycle of this kind was first suggested by Carnot. It consists of the following four processes (Fig. 165).

(*a*) During the *isothermal expansion* at the temperature T_1 from the state A to state B the work done equals the area $ABba$ and a certain amount of heat, H_1, is absorbed by the expanding gas.

(*b*) During *adiabatic expansion* from B to C the work done equals area $BCcb$, no heat is absorbed and the temperature drops to T_2.

(*c*) Now let the gas be *compressed isothermally* from C to D. External work must be spent equal to area $DCcd$ and an amount

of heat, H_2, is given out by the gas at the lower temperature T_2.

(*d*) Finally let the body return to its original state by an *adiabatic compression;* the external work to be spent equals area *ADda*.

By performing this cycle the following lasting results are obtained:

1. A quantity of heat, H_1 has been absorbed at a temperature T_1;

2. A quantity of heat, H_2, has been given out at the temperature T_2;

3. A certain amount of mechanical work represented by the area included between the two isothermal and two adiabatic lines:

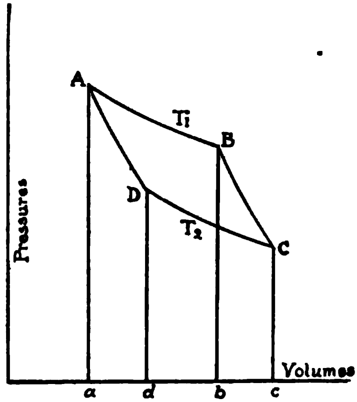


FIG. 165.

$$ABCD = ABba + BCcb - DCcd - ADda$$

has been performed and must have been obtained, according to the first law of thermodynamics, by a disappearance of a quantity of heat equal to $H_1 - H_2$.

The body performing the cycle is in the end in exactly the same condition as before; it has been only the agent by means of which the above results were obtained and is therefore called "the working substance."

Carnot also imagined an engine in which this cycle may be used for the continued production of work from heat (Fig. 166).

Let the working substance, for example, a gas, be enclosed in a cylinder with walls impermeable to heat, in which a frictionless piston, also non-conducting, moves; but let the bottom of the cylinder be perfectly conducting. Besides this we need a source of heat, the "heater" at a higher temperature T_1, a "refrigerator" at the lower temperature T_2 both perfectly conducting, and a stand *S*, of non-conducting material. The four parts of Carnot's cycle may then be realized by the following manipulations.

(*a*) The cylinder at the temperature T_1 is placed in contact with the *heater* and the gas expanded isothermally.

(*b*) The cylinder is transferred to the non-conducting *stand* and the expansion continued adiabatically until the temperature has fallen to T_2.

(*c*) The cylinder is placed in contact with the *refrigerator* and compressed isothermally.

(*d*) The cylinder is again placed upon the *stand* and compressed adiabatically until the temperature has risen to T_1.

These four operations might just as well have been made in the reverse order; for example we might have started with the cylinder in contact with the refrigerator, at the temperature T_2 and under

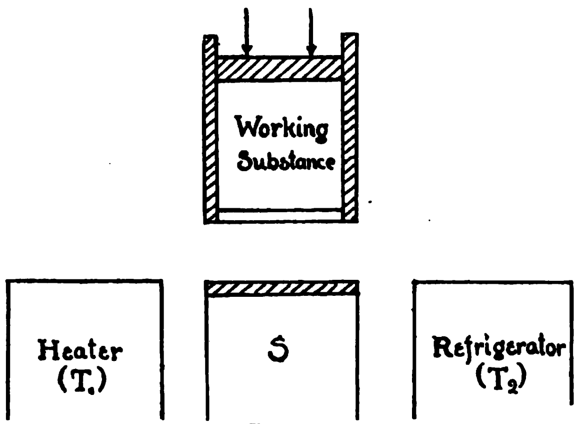


Fig. 166.

a pressure corresponding to the point D of figure 165. Then let the gas expand isothermally until the point C is reached. A quantity of heat equal to H_2 will be absorbed. The next operation would be an adiabatic compression, with the cylinder upon the non-conducting stand. Then the cylinder is placed upon the heater and the gas compressed isothermally. The quantity of heat equal to H_1 is given to the heater. Finally the cylinder is placed again upon the stand S and the gas expanded adiabatically until its temperature has fallen to T_2. The result of this reversed process is, that heat is transferred from a lower temperature to higher, while at the same time mechanical work is transformed into heat, *i. e.*,

the effect of the engine is reversed with the operations. For this reason Carnot's hypothetical engine is a "reversible" one.

325. Reversible and Irreversible Processes. While the first law of thermodynamics expresses simply the equivalence of heat and mechanical work, it says nothing about the direction in which a natural phenomenon takes place. When a bullet strikes a target, heat is produced equivalent to the kinetic energy destroyed. We know that this heat is unable directly to reverse the phenomenon.

Processes can be imagined in which nature seems to have no preference as to the direction of a transformation. In the case of a simple pendulum moving without friction, or in vibrations of a perfectly elastic medium the change of kinetic into potential energy is seemingly as complete as the reverse process. But in other transformations the "natural" direction is perfectly definite. Such natural processes are, for example,

(*a*) Transference of heat from a hotter to a cooler body.

(*b*) Production of heat by destruction of mechanical work.

(*c*) Expansion of gases.

We shall call the processes taking place in the opposite direction from the natural artificial processes, for example,

(*a*) Transference of heat from a cooler to a warmer body.

(*b*) Production of work from heat.

(*c*) Compression of gases.

These cannot take place alone, but must always be accompanied by a natural process. For example (*a*) by reversing Carnot's cycle heat may be transferred from the refrigerator to the reservoir at higher temperature, but at the same time a transformation of work into heat takes place.

(*b*) Mechanical work can be obtained from heat, as in the direct Carnot cycle, but not without a transference of heat from a higher to a lower temperature. Heat is also transformed into work during the expansion of gases, but the latter is a natural process.

(*c*) A compression of a gas is impossible without mechanical work being transformed into heat, which latter appears in the gas.

A natural process can take place either alone or accompanied by an artificial one. Only in the latter case can the process be reversed. Thus in Carnot's cycle we have a combination of heat transference and production of work. Reversing it, the heat can

be transferred in the opposite direction by retransforming the work, obtained by the first operation, into heat. *In every such cycle there must be a combination of a natural and artificial process at each step of the cycle.*

There are, however, phenomena in nature, in which only the natural process takes place, without an equivalent artificial one, for example,

(*a*) Conduction of heat,

(*b*) Production of heat by friction,

(*c*) Expansion of a gas into a vacuum.

None of these processes can be reversed without a lasting change being produced in the system. Let, for example, heat be conducted directly from the heater to the refrigerator in Carnot's engine. To reverse this process the engine must be worked backwards, and, after the heat is restored to the heater, mechanical work from the outside has disappeared and the whole system taking part in the cycle is not in the original condition.

Such irreversible processes as conduction of heat, friction and free expansion of gases will make any process of which they form a part irreversible, and they appear to a greater or less extent in all natural phenomena.

326. The Second Law of Thermodynamics. The second law of thermodynamics is an attempt to formulate in a single rule the complicated relations discussed in the last article, and thus complement the first law, by indicating the direction in which a natural phenomenon will proceed. According to the point of view it can be stated in different forms. Clausius worded it thus: "Heat cannot of itself pass from a colder to a warmer body." This "of itself" must be understood as "without a natural process accompanying it."

Lord Kelvin formulated the same principle thus: "It is impossible, by means of inanimate material agencies to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest surrounding objects." Mere cooling would be a natural process without transformation of heat into some other form of energy.

These statements, being negative, are not based upon experimental facts. As soon as an experiment could be made showing

that one of the processes, mentioned in the last article as irreversible, can be reversed, the second law with all its conclusions would fall; but that seems highly improbable.

327. Efficiency of Carnot's Cycle. By efficiency we understand here the ratio of the quantity of heat transformed into work to the quantity of heat absorbed by the working substance at the higher temperature, or

$$e = \frac{H_1 - H_2}{H_1} \tag{122}$$

Carnot derived the important principle, that *no engine can be more efficient than a reversible engine and the efficiency of all reversible engines working between the same two temperatures is the same.*

For, let us suppose an engine, A, to have a greater efficiency than a reversible engine, C, such as Carnot's engine, and to be coupled with C, the latter working backwards. Let A absorb a quantity of heat, H_1, at the higher temperature and give out H_2 at the lower temperature, while C absorbs H_2' at the lower and returns H_1' to the higher temperature. Then $H_1 - H_2$ is the heat converted into work and $H_1' - H_2'$ the work converted into heat during one cycle of each engine. Let A produce only $1/n$ as much work in one cycle as C. If now A be coupled to C in such a way that it makes n cycles to each one of the C's the work done by A is equal to that done on C, so that no work is left over. Since the total work performed by each engine must equal the heat transformed

$$n(H_1 - H_2) = H_1' - H_2' \tag{123}$$

The efficiency of engine A is

$$\frac{H_1 - H_2}{H_1} = \frac{n(H_1 - H_2)}{nH_1} \tag{124}$$

Hence according to our supposition

$$\frac{n(H_1 - H_2)}{nH_1} > \frac{H_1' - H_2'}{H_1'} \tag{125}$$

But since the numerators of these fractions are equal by (123) H_1' would be larger than nH_1, or engine C would return more

heat to the reservoir at the higher temperature than is taken from it by A. As the only result of the working of the two engines heat would be transferred continuously from the cooler to the hotter reservoir, which is contrary to the second law of thermodynamics (see § 326). The assumption that A has a larger efficiency than C is therefore inadmissible.

Hence no engine, reversible or not, can have a higher efficiency than a reversible engine and therefore all reversible engines are equally efficient.

The efficiency of a reversible engine is therefore independent as well of the nature of the working substance as of the quantity of heat absorbed at the higher temperature, and *is only a function of the temperatures between which the engine works.*

328. Efficiency and Temperatures in a Reversible Cycle. The equations for isothermal and adiabatic expansion of a perfect gas give:

$$P_1 V_1 = P_2 V_2 \quad \text{or} \quad \frac{P_2}{P_1} = \frac{V_1}{V_2} \tag{126}$$

$$P_3 V_3 = P_4 V_4 \quad \text{or} \quad \frac{P_3}{P_4} = \frac{V_4}{V_3} \tag{127}$$

$$P_2 V_2^\gamma = P_3 V_3^\gamma \tag{128}$$

$$P_1 V_1^\gamma = P_4 V_4^\gamma \tag{129}$$

Dividing (128) by (129) and substituting (126) and (127)

$$\frac{V_2^{\gamma-1}}{V_1^{\gamma-1}} = \frac{V_3^{\gamma-1}}{V_4^{\gamma-1}} \tag{130}$$

or

$$\frac{V_2}{V_1} = \frac{V_3}{V_4} \tag{131}$$

The four volumes are therefore not independent of each other.

The amount of heat absorbed during the first operation is

$$H_1 = \int_{V_1}^{V_2} P\,dV = PV\int_{V_1}^{V_2} \frac{dV}{V} = P_1 V_1 \int_{V_1}^{V_2} \frac{dV}{V} = P_1 V_1 \log \frac{V_2}{V_1} \tag{132}$$

similarly for the third operation:

$$H_2 = -\int_{V_3}^{V_4} P\,dV = P_3 V_3 \log \frac{V_3}{V_4} \tag{133}$$

Dividing (132) by (133) and considering equation (131) and the gas law

$$\frac{H_1}{H_2} = \frac{P_1 V_1}{P_3 V_3} = \frac{R_1 T_1}{R_1 T_2} = \frac{T_1}{T_2} \tag{134}$$

Hence, if we measure the temperature by the gas law, *i. e.*, take a perfect gas as thermometric substance, the quantities of heat absorbed and given out by the working substances are proportional to the absolute temperatures, and

$$e = \frac{H_1 - H_2}{H_1} = \frac{T_1 - T_2}{T_1} \tag{135}$$

329. The Thermodynamic Scale of Temperature. Since the efficiency of a reversible engine working between two given temperatures depends on these temperatures only, Lord Kelvin suggested that this property might be used for the construction of a truly "absolute" scale of temperature, *i. e.*, one independent of any thermometric substance. We may thus agree that, counting from a given temperature, θ_1, of the warmer reservoir, the interval between this temperature and that of the refrigerator, θ_2, shall be proportional to the efficiency of a reversible engine working between the two temperatures.

Since θ_1 is a constant, the efficiency is expressed by

$$e = A(\theta_1 - \theta_2) \tag{136}$$

The centigrade principle should be chosen to fix the value of the unit of the scale. Thus if θ_1 represents an isothermal at the temperature of boiling water and θ_2 the temperature of melting ice, the 99 isothermals between the two should be located so that the area included between two consecutive isothermals and a given pair of adiabatics equals 1/100th of the area between the isothermals at θ_1 and θ_2 and the same adiabatics (§ 241).

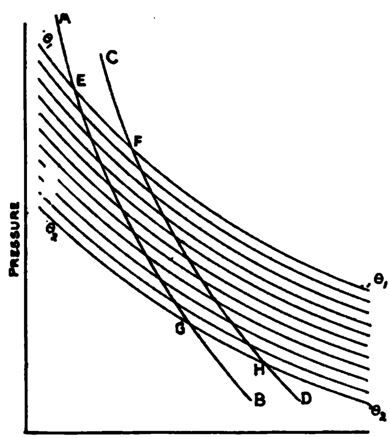


FIG. 167.

Supposing that we can lower the temperature of the refrigerator sufficiently so that all heat absorbed is converted into work, this temperature must be *the absolute zero.* For, if still

lower temperatures existed, we could, by selecting this as the temperature of the refrigerator get more work out of the absorbed heat than its mechanical equivalent; but this contradicts the first law of thermodynamics. If the refrigerator is at absolute zero, the efficiency would be unity, and

$$1 = A\theta_1 \quad \text{or} \quad A = 1/\theta_1 \tag{137}$$

With any other temperature of the refrigerator

$$e = \frac{H_1 - H_2}{H_1} = A(\theta_1 - \theta_2) = \frac{\theta_1 - \theta_2}{\theta_1} \tag{138}$$

This gives

$$\frac{H_1}{H_2} = \frac{\theta_1}{\theta_2} \tag{139}$$

Hence the quantities of heat absorbed and given out by a reversible engine are proportional to the absolute temperatures on the thermodynamic scale.

Equation (139) is identical with (134). The thermodynamic scale would therefore agree with one obtained with a perfect gas as thermometric substance, and differ the more from an ordinary gas thermometer the more the gas employed differs from a perfect gas. The corrections to be applied to a given gas scale can be calculated if the deviations of the gas from the gas law are known. These have been determined by the porous plug experiment (§ 304). Since hydrogen differs but little in its properties from a perfect gas the hydrogen scale must closely agree with the thermodynamic scale: in fact Callendar calculated that between $-50°$ and $150°$ the differences amount to less than $0.001°$, and even at $1000°$ the readings on the hydrogen scale are only $0.044°$ lower than those on the thermodynamic scale. Buckingham more recently found practically the same.

While in the earlier part of the book (§ 249) the expression "absolute temperature" was defined as the temperature counted from the zero of the gas thermometer, it should be defined as the temperature measured from the absolute zero of the thermodynamic scale. This ambiguity leads, however, to no serious error on account of the close agreement of the scales.

330. Variation of the Freezing Point with Pressure. Let one gram of water be frozen under atmospheric pressure of $0° = T_2$. The amount of work done is

$$W = P(v_i - v_w) \tag{140}$$

where v_i and v_w are the specific volumes of ice and water respectively. Their difference is 0.0907 cm.^{3}, $P = 1013000$ dynes per cm.^{2}, therefore

$$W = 91{,}880 \text{ ergs}$$

Now let the pressure be changed to zero, while the temperature rises to $T_1 = T_2 + \Delta T$, the melting point of ice under zero pressure. (See also § 294). No further work is done since the volume of ice does not change appreciably, and since ΔT is a very small fraction of a degree, also the thermal effects are negligible. At T_1 the ice is melted again. The external work is zero, since P equals zero and the amount of heat absorbed is $L = 80$ cal. $= 335 \times 10^7$ ergs. Finally the pressure is increased to one atmosphere while the temperature sinks to T_2.

The substance has thus been carried through a complete cycle which is evidently reversible, since all the operations could have been performed in the opposite direction. Equations (122) and (138) give

$$e = \frac{W}{L} = \frac{\theta_1 - \theta_2}{\theta_1} = \frac{T_1 - T_2}{T_1} \tag{141}$$

Since T_1 does not differ much from $T_2 = 273°$ we obtain for the change of the freezing point with a change in pressure of one atmosphere

$$T_1 - T_2 = \frac{273 \cdot 91880}{335 \cdot 10^7} = 0.0075° \tag{142}$$

a value which agrees well with the experimental result (§ 278).

331. Reversible Cycles in General. In § 324 we have only considered a reversible cycle in which the heat was all absorbed at the same temperature and given out at another constant temperature. The treatment may be generalized so as to include reversible cycles in which heat may be absorbed or given out at any temperature. Let such a cycle, performed by an engine, be represented by the closed curve of Fig. 168, of which the ordinate represent pressures and the abscissas volumes of the working substance. Any part of the curve taken at random is neither an isothermal nor an adiabatic line; but the whole cycle may be divided, as follows, into a large number of small Carnot cycles.

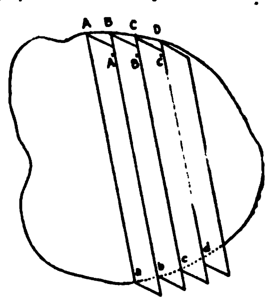


FIG. 168.

Draw isothermal and adiabatic lines through neighboring points $A, B, C, \cdots a, b, c$. If the points are chosen sufficiently close together the heat absorbed along the curve AD will equal that absorbed along $AA'BB'CC'D$. Similar reasoning would apply to the heat given out at the lower boundary.

The heat absorbed and given out by the arbitrary reversible cycle is therefore the same as for the sum of the small Carnot's cycles of which the former may be assumed to be made up. Since in the limit the area of the cycle becomes equal to the sum of the areas of the smaller ones, the

work performed is the same as if we had passed through the large number of small Carnot's cycles. Since for each of the small cycles equation (139) holds,

$$\frac{H_1}{\theta_2} - \frac{H_2}{\theta_1} = 0 \tag{143}$$

and consequently for the general cycle

$$\Sigma\frac{H_1}{\theta_2} - \Sigma\frac{H_1}{\theta_2} = 0 \tag{144}$$

Taking H with the proper sign, positive if the heat is added to a body, negative if it is lost, the last equation can be written

$$\Sigma\frac{H}{\theta} = 0 \qquad \text{or} \qquad \int\frac{dH}{\theta} = 0 \tag{145}$$

Equation (145) shows that dH/θ is a perfect differential, say dS. The function S is called *entropy*. The entropy of the substance which has been put through a reversible cycle is unchanged; that of the heater is diminished and that of the refrigerator is increased by the same amount. Therefore: *In any reversible cycle the entropy of the system remains constant.*

332. Increase of Entropy. The absolute value of the entropy of a body is unknown, but we have to deal only with entropy changes and may take any standard state as reference point. If a reversible change occurs in which heat is neither absorbed nor given out, the entropy remains constant; in this case an adiabatic change is also an isentropic one.

For irreversible processes the entropy of the system always increases. In heat conduction between two isolated bodies, for example, heat is lost by the hotter body at θ_1 and gained by the cooler body at θ_2. The loss of entropy dH/θ_1 is therefore smaller than the gain dH/θ_2. But since the former is negative, we have for the change of entropy of the two bodies, between which heat exchange by conduction takes place

$$dS = dH\left(\frac{1}{T_2} - \frac{1}{T_1}\right) > 0 \tag{146}$$

The same proposition can be proved for other irreversible processes.

Since all natural phenomena are accompanied by some irreversible process we may state: During any transformation in nature the entropy tends to increase. Such irreversible processes as conduction and radiation produce an equalization of temperature and the entropy tends towards a maximum which is reached when a uniform temperature exists throughout and a transformation of heat into mechanical work has become impossible.

The second law of thermodynamics becomes thus the principle of the increase in entropy which, with the law of conservation of energy forms the foundation of modern Physics. In fact the increase in entropy can be taken as a measure for the irreversibility of a process or for the preference nature has for it.

333. Efficiency of an Irreversible Cycle. Let a process which is only partly reversible be carried through a cycle between the temperatures θ_1 and θ_2. According to the last article

$$\frac{H_1}{\theta_1} < \frac{H_2}{\theta_2} \tag{147}$$

Hence

$$\frac{H_1 - H_2}{H_1} < \frac{\theta_1 - \theta_2}{\theta_1} \tag{148}$$

But $(H_1 - H_2)/H_1$ is the efficiency of the irreversible cycle and $(\theta_1 - \theta_2)/\theta_1$ that of a reversible cycle. The efficiency of an irreversible engine is therefore always smaller than that of a reversible engine working between the same temperatures. The result can easily be generalized in the same way as in § 331.

334. The Reciprocating Steam Engine. The most important method for transforming heat into mechanical work is by means of a steam engine, in which steam plays the rôle of the working substance.

In this engine we may distinguish the following partial operations:

(*a*) Steam coming from a boiler produces external work by moving a piston against pressure. The process takes place practically under constant pressure, the pressure of the steam.

(*b*) The communication with the boiler is shut off but the steam still expands against external pressure. This corresponds to the adiabatic expansion in Carnot's cycle, but in fact is nearly an isothermal process due to the conduction from the walls and the effects of condensation.

(*c*) The steam, partly condensed, is connected with the outside and the volume reduced, while the steam escapes into the air or a condenser kept at a lower pressure than the air. There the steam is condensed to water. This is also a process under constant pressure.

(*d*) Communication with the outside is shut off and the remaining steam compressed with a rise of temperature to the original pressure.

To make the cycle complete for the working substance, we may suppose that the water is returned from the condenser to the boiler to be transformed again into steam.

The schematic sketch, Fig. 169, will illustrate the working of the engine. Suppose the piston *P* in such a position that steam is

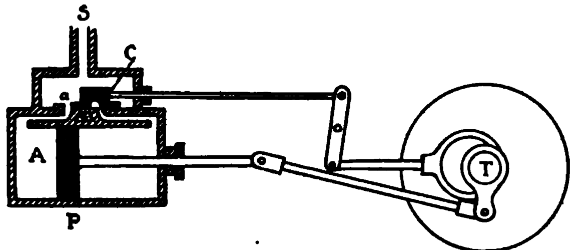


FIG. 169.

admitted through the steam port, *a*, into the cylinder *A*. The piston is moved towards the right, producing a rotation of the shaft *T*. A slide valve *C* is connected with the shaft so as to be moved towards the left by the rotation of *T*; it soon shuts off the steam. The motion of the piston continues, however, increasing the volume *A* and decreasing the pressure. Shortly before the piston reaches its extreme position at the right, the valve *C* has moved over far enough to connect *A* with the exhaust port *e* thus reducing the pressure to that in the condenser. Now the decrease in volume begins since the piston is pushed in the opposite direction by steam which is admitted by the slide valve to the other side of the cylinder. Finally the slide valve, which moves now back towards the right, shuts *A* off again and the remaining steam is compressed to the original pressure by the continued motion of the cylinder to the left, when the valve reëstablishes communication with the steam port and the cycle begins anew. Engines in which a piston moves backward and forward are called reciprocating engines.

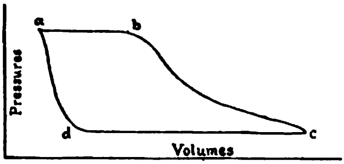


FIG. 170.

By a proper mechanism the relation between the volume and the pressure of the steam can be traced simultaneously upon a card. This diagram is known as the indicator card, Fig. 170.

It shows distinctly in the four

parts *ab*, *bc*, *cd*, and *da* of the curve the four operations mentioned above. This cycle differs considerably from a Carnot's cycle though bearing a general resemblance to it.

335. The Efficiency of a Steam Engine is considerably below that of a reversible cycle. To increase its efficiency the step from the high boiler pressure to that of the exhaust is usually divided into two, three or four steps (compound, triple or quadruple expansion engine) since higher initial pressures and temperatures can be used in such cases.

EFFICIENCY OF STEAM ENGINES.

	Temperature.		Efficiency.	Efficiency of Reversible Cycle.
	t_1	t_2		
Willan's Engine (non condensing)	164°	101.5°	10.4%	14.5%
Levitt Pumping Engine (compound)	181.6	37.7	19	31.7
Levitt Pumping Engine (triple expansion)	191.9	46.7	20.8	31.8
Nordberg Engine (quadruple expansion)	206.35	43.1	25.5	34.0

In § 268 the heat of combustion of one gram of coal was given as about 8000 calories $= 33.5 \times 10^{10}$ ergs. A part of this is lost before it reaches the boiler. The greatest efficiency of a steam engine obtained so far is 25 per cent. Therefore under the most favorable conditions one gram of coal will yield about 8.5×10^{10} ergs, neglecting the loss from the furnace to boiler.

Expressed in English units the calculation would be thus:

One pound of coal produces at the boiler about 10,000 British Thermal Units; one B. T. U. equals 778 foot pounds. One horse power for one hour $= 550 \times 3{,}600 = 198 \times 10^4$ foot pounds.

With highest efficiency of the engine one pound of coal furnishes about one horse power for one hour; but with non-condensing engines two to three pounds of coal are needed to perform this amount of mechanical work.

336. The Steam Turbine. In the steam turbine jets of steam are directed through properly constructed nozzles against vanes mounted on the periphery of a revolving disk or drum, *i. e.*, the potential energy of the steam is converted into kinetic energy

before doing useful work. The steam striking the vanes exerts a considerable force upon them (see § 204) and thus produces rotation of the wheel.

Fig. 171 is a diagram illustrating the cross-section of a turbine, in which the interior surface of the case is covered with stationary blades (S) alternating with those on the drum (M). The steam passing alternately between the stationary and movable blades finally leaves the machine, after it has given up its kinetic energy to the revolving part. Such turbines are said to be of the "impulse type." The expansion of the steam may take place either in a single stage, in which case the peripheral velocity of the drum must be very high, from 700 to 1,400 feet per second, or the expansion is divided between a number of stages, each one exhausting through properly proportioned nozzles into the next succeeding stage (multistage turbines). In the latter type the angular velocity is much reduced.

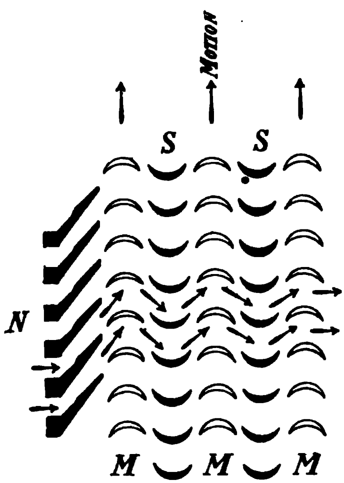


FIG. 171.

In the "reaction type," of which the Parsons and Allis-Chalmers turbines are good examples, the expansion of the steam takes place partly before doing work upon the wheel and partly within the turbine itself. In these machines the steam passes through a large number of rows of blades alternately fixed and movable. The shaft is coaxial with the cylinder through which the steam flows and the blades are placed upon the shaft in the same way as the blades of a windmill; the cross-section of the cylinder enlarges in the direction in which the steam flows, allowing an additional expansion of the steam inside the machine.

The principal advantages of the steam turbines over the reciprocating engine are economy of space, freedom from vibration and uniform and high angular velocity. They are especially adapted

for driving alternate current dynamos, centrifugal pumps, etc., where high angular velocity is required, and in steamships on account of their smooth running. The difference in efficiency between the steam turbine and the best types of reciprocating engines is small.

337. The Internal Combustion Engine. In these engines the fuel supplying the energy is burnt within the engine proper. The engine consists of a cylinder in which a piston, connected in the ordinary manner to the rotating part of the engine, moves back and fourth. When the piston is nearest the cylinder head there is still some space left between the walls. This clearance is called the explosion chamber. In the clearance walls are two valves, through one of which, the "inlet valve" (*i* in Fig. 172), the fuel supply or charge is admitted; while the products of combustion are expelled through the other, the "exhaust valve" (*e*). In the most common type of these engines four strokes of the piston are required to complete the series of operations taking place in the cylinder, the so-called "Otto cycle" (see Fig. 172). In (1), the *outstroke* or suction stroke, the piston moves forward while the inlet valve is opened. The explosive mixture fills the cylinder. At the end of this stroke the inlet valve closes and during (2), the *compression stroke*, the piston moves back and compresses the gas within the clearance space. At the end of this stroke the charge is exploded, usually by an electric spark. The explosion produces a very high temperature of the gaseous mixture and a corresponding high pressure which (3) pushes the piston forward again, *expansion stroke*. At the end of this stroke the exhaust valve opens and during (4), the *instroke*, the burnt gases are expelled and the engine is again in the original state.

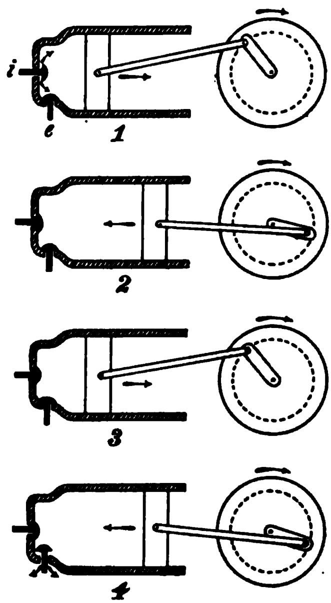


FIG. 172.

In some engines, especially in those for marine use, only two strokes are required instead of four (two cycle engines). In others two cylinders are placed in tandem, thus giving a double acting effect as in the reciprocating steam engine.

Gases or vapors of oil, gasoline or alcohol are used as fuel in the internal combustion engine. The preparation of the fuel requires in general three steps: vaporization, mixing with air, and adjustment to proper proportions. This is either done in separate apparatus, called vaporizers and carbureters, or, the vapors of the more easily vaporizing liquids are formed directly in the combustion chamber by dropping it upon a heated plate.

The internal combustion engine, being its own generator of heat, and, therefore, not being handicapped by a separate generator, is likely to be much more economical than the engines discussed in the two previous sections. It is used extensively in automobiles, motor boats, etc., where it is necessary to carry the fuel along with the engine. Another advantage of the small gas engine is the fact that it can be started in a very short time.

338. Dimensional Formulae. Temperature cannot be expressed in terms of mass, length and time; it forms a new fundamental unit, in addition to those used in mechanics (§ 151–155) and may be called a secondary fundamental unit. It enters into the dimensions of many thermal quantities and is represented for this purpose by $[\theta]$.

Heat has been shown to be energy (§ 317); its dimensional formula must thus be $[ML^2T^{-2}]$, while the mechanical equivalent of heat, J, whose dimensional formula according to equation 78 is $[J] = [W]/[H]$, has no dimensions and is therefore a pure number.

The dimensional formulae for all other thermal quantities are easily derived from the equations defining them.

a. Temperature coefficient from equations 2, 9 or 14: $[a] = [\theta^{-1}]$.

b. The gas constant from equation 39: $[R] = [L^2T^{-2}\theta^{-1}]$.

c. Heat capacity of a body from equation 41: $[C] = [ML^2T^{-2}\theta^{-1}]$.

d. Heat capacity of a substance becomes independent of the mass by equation 43: $[c] = [L^2T^{-2}\theta^{-1}]$.

e. Specific Heat, the ratio of the heat capacities of two substances, is a pure number.

f. Temperature gradient $= (t_2 - t_1)/l$ (p. 216). $[\text{gradient}] = [L^{-1}\theta]$.

g. The coefficient of thermal conductivity is given by equation 68. $[k] = [MLT^{-3}\theta^{-1}]$.

h. Heat of fusion and heat of vaporization of a substance from pages 225 and 233: $[L] = [L^2T^{-2}]$.

REFERENCES.

TYNDALL'S *Heat as a Mode of Motion* contains Tyndall's famous popular lectures on heat, illustrated by interesting experiments.

PRESTON'S *Theory of Heat* is a textbook for somewhat advanced students. Detailed description of classical methods and experiments.

POYNTING AND THOMSON'S *Heat* is a textbook with emphasis placed upon the mathematical treatment.

LECHATELIER BOUDOUARD'S *High Temperature Measurements* (transl. by Burgess). Full description of the various methods for the measurement of high temperatures.

DAVIS' *Elementary Meteorology.* An excellent text of this subject.

HARDIN'S *Liquefaction of Gases.* Contains the history of this subject and full description of the various methods and apparatus used.

PLANCK'S *Treatise on Thermodynamics* (transl. by Ogg) is an advanced treatise especially suited for students of physics and chemistry.

SPANGLER, GREENE AND MARSHALL'S *Elements of Steam Engineering* describes the different types of engines and boilers in use and illustrates the construction of their various parts in detail.

MOYER'S *Steam Turbine* deals with the construction and operation of the steam turbine as well as the theory and design.

JONES' *Gas Engine* gives the history and description of different types and treats at considerable length of the care and operation of engines.

Harper's Scientific Memoirs, edited by Ames, is a collection of classical papers either as reprints or English translations and each paper is preceded by a short biographical sketch of the author. Of these books the following are of interest:

Ames' *The Free Expansion of Gases* contains the following papers:

- Gay-Lussac, First attempt to determine the changes in temperature which gases experience owing to changes of density, etc. (1807).
- Joule, On the changes of temperature produced by the rarefaction or condensation of air (1845).
- Thomson (Kelvin) and Joule, On the thermal effects of elastic fluids (1853–1862).

Magie's *The Second Law of Thermodynamics* contains:

- Carnot, Reflections on the motive power of heat (1824).
- Clausius, On the motive power of heat, etc. (1850).
- Kelvin, The dynamical theory of heat (1851).

Also the following papers, being either descriptions of classical experiments or good summaries of particular fields:

Day and Clement, Am. Jour. of Sci., 26, p. 405, 1908, contains a description of a very carefully constructed standard gas thermometer and temperature measurements made with it.

Waidner and Dickinson, Bull. Bur. of Standards 3, p. 663, 1907, contains a description of thermometers, methods of intercomparison, corrections to be applied, etc., for the interval from 0° to 100°.

Andrews, Phil. Trans., 1869, p. 575, contains the experiment on the liquefaction of carbon dioxide; discovery of the critical point.

Joule, Phil. Trans., 1850, p. 61 and 1878, p. 365, contains the experimental determination of the mechanical equivalent of heat.

Rowland, Proc. Am. Ac. Sci. (new series), 7, p. 75, 1879, contains Rowland's determination of the mechanical equivalent of heat.

Barnes, Trans. Intern. Electr. Congress, St. Louis, I, p. 53, 1905, contains a good summary of the determinations of the mechanical equivalent of heat by electrical methods.

Waidner and Burgess, Bull. Bur. of Standards, 1, p. 189, gives a full description of the methods and apparatus employed in pyrometry.

PROBLEMS.

Thermometry.

1. Reduce the following readings on a Fahrenheit thermometer to the corresponding readings on a centigrade thermometer: $+1500°$, $+95°$, $+14°$, $-49°$.

Ans. $+815.5°$ C.; $+35°$ C.; $-10°$ C.; $-45°$ C.

2. Reduce the following readings on a centigrade thermometer to the corresponding readings on a Fahrenheit thermometer: $+1500°$, $+230°$, $-15°$, $-273°$.

Ans. $+2732°$ F.; $+446°$ F.; $+5°$ F.; $-459.4°$ F.

3. At what temperature do Fahrenheit and centigrade thermometers read the same? At what temperatures does (*a*) the Fahrenheit thermometer read twice, and (*b*) three times as much as a centigrade thermometer?

Ans. $-40°$; (*a*) $+320°$ F.; (*b*) $+80°$ F.

Expansion.

4. Railroad rails, each 30 feet long, are laid at a temperature of 60° Fahrenheit. How large a gap must be left between the ends if the highest temperature to be allowed for is 110° Fahrenheit? Ans. 0.117 in.

5. In a gridiron pendulum (§ 251) the distance from the center of suspension to the center of oscillation is 99.3 cm. Supposing that all parts of the pendulum except the two vertical rods on either side of the central one are of iron, how long must these two rods be if made of zinc? Why cannot brass be used instead of zinc in this type of pendulum?

Ans. 74.0 cms.

6. How large a relative error is made if the moment of inertia of a brass cylinder, calculated from measurements at 10°, is used in finding τ (§ 119) for torsional vibrations at 50° without being corrected for temperature? Ans. 0.14 per cent.

7. A clock with a seconds pendulum of steel is correct at 25°. How many seconds will it gain in a day if the temperature is 0°?

Ans. 12.6 sec.

8. A liter flask of Jena glass 16^{III} is calibrated at 0°. The cylindrical neck has a diameter of 1.5 cm. How far above the liter mark will the meniscus be if the flask is filled with one kilogram of water at 30°?

Ans. 2.08 cms.

9. Reduce to 0° a barometer reading of 74.5 cm. taken at 19°, considering both the expansion of the mercury and the brass scale.

Ans. 74.27 cms.

10. A piece of brass weighs 17 grams in air and 15 grams in water of 10°. How much will it weigh in water of 30°? Ans. 15.007 g.

11. A mass of a perfect gas at 20° occupies 200 c.c. Find its volume at 90°, the pressure remaining constant. Ans. 247.8 c.c.

12. 20 c.c. of hydrogen gas is collected in a closed tube over mercury. The barometer reading corrected for temperature is 74.5 cm. and the temperature of the gas 24°. The mercury in the tube stands 12 cm. above the mercury in the dish in which the tube is inverted. Reduce the volume of the gas to 0° and normal atmospheric pressure. Ans. 15.12 c.c.

Calorimetry.

13. Calculate the water-equivalent of unit volume of mercury and of glass, the density of the latter being 2.5 g. per c.c. Show that the water-equivalent of the immersed part of a thermometer may be taken without an appreciable error as numerically equal to 0.47 of its volume. Ans. 0.449 and 0.498 g. of water per c.c.

14. 300 g. of copper, heated to 99.4° are dropped into 400 g. of water contained in a copper cup whose mass is 90 g. The temperature of the calorimeter is raised from 20° to 25.1°. Find the specific heat of copper.
Ans. 0.0935

15. If in the last problem alcohol is used instead of water, how large would the rise in temperature be and what would be the specific heat of copper referred to alcohol as a standard? Ans. (*a*) 8°; (*b*) 0.155.

16. How much heat, expressed in B.T.U., is produced by the burning of one pound of anthracite coal? Ans. 14120 B.T.U.
one pound of anthracite coal?

17. A lake having a surface area of 9000 square meters is covered by a sheet of ice 5 cm. thick. How much heat would pass through the ice in two hours if the temperature of the air is — 10° and the ice does not increase appreciably in thickness?
Ans. 648×10^7 cal.

Conduction of Heat.

18. How much coal must be burned to compensate for the loss due to conduction of heat for one day through a glass window, 4 mm. thick and having an area of two square meters, supposing that the air in the room next to the glass is at 25° and the outside air at — 10°? Why is this amount much larger than that actually needed to keep a room at 25° when the temperature ouside is — 10°? Ans. 28.9 Kg. of anthracite.

19. How much ice at 0° will be melted by 50 g. of copper which has been heated in to 200°? Ans. 11.75 gr.

Change of State.

20. How much heat is necessary to change 50 g. of ice at — 10° to steam at 150°? Ans. 37200 cal.

21. 100 g. of ice at — 10° are dropped into a nickel calorimeter whose mass is 100 g. and which contains 500 g. of water. The temperature of the calorimeter is lowered from 30° to 11.9°. How large is the heat of fusion of water? Ans. 75.52 cal./gr.

22. One side of a copper plate 1.5 cm. thick and 10 cm.^{2} in cross-section is kept at 100° by means of steam and the other is in contact with melting ice. 4.5 kilograms of ice are melted in 10 minutes by the heat conducted through the plate. Find the coefficient of thermal conductivity of copper.
Ans. 0.9 units.

23. Heat is supplied at a constant rate to a block of tin. It is found that the temperature of the tin increases 3° in 2 seconds. As soon as the melting point is reached the temperature remains constant for 160 seconds, *i. e.*, until all the tin is melted. How large is the heat of fusion of tin? Specific heat of tin = 0.055. Ans. 13.2 cal./g.

24. How large a portion of water undercooled to — 12° will freeze when crystallization takes place. Disregard the water equivalent of the vessel. Ans. 15 per cent.

25. How much would the air in a room, $6 \times 5 \times 3$ meters, be warmed by the condensation alone of one kilogram of steam in the radiator?
Ans. 19.4°.

26. How much heat is absorbed when one kilogram of liquid air is boiled under atmospheric pressure and subsequently heated to 20°? Compare this amount with the heat absorbed by one kilogram of ice melted at 0° and subsequently heated to 20°. Boiling point of liquid air = — 190°, heat of vaporization of air = 50 cal. per gram.
Ans. (*a*) 99770 cal.; (*b*) 100000 cal.

27. Draw on accurate cross-section paper the vapor tension curve for water and determine from the curve at what temperature water would boil under a pressure of (*a*) 10 cm. of mercury, (*b*) 40 cm. of mercury. (*c*) two atmospheres. Ans. (*a*) 51.7°; (*b*) 83.0°; (*c*) 121°.

28. What is the relative humidity of air at 20° if the dew point is found to be 10°? Ans. 52.7 per cent.

29. How much work would be performed if the total heat of combustion of one kilogram of anthracite could be transformed into work? Ans. 9.131 kilo-watt hours.

Thermodynamics.

30. The Niagara falls are 160 feet high. How much warmer should the water be at the foot of the falls than at the top. Disregard effect of evaporation. Ans. 0.206° F.

31. When a street car weighing 4000 kilograms and having a speed of 20 kilometers per hour is stopped by the brakes how much heat is produced? Ans. 14800 cal.

32. What must be the speed of a lead bullet if on striking a target its temperature is raised from 27° to the melting point? Assume that all heat produced serves to heat the bullet. Ans. 279 m./sec.

33. How much external work is done by 1 kilogram of water when it freezes at 0° under atmospheric pressure? What would be the heat of freezes at 0° under atmospheric pressure? What would be the change in the heat of fusion of water if this amount of energy were not included? Ans. (*a*) 9.27×10^7 ergs; (*b*) 0.002 cal./g.

34. How much external work is done by the transformation at 100° of 1 kilogram of water to steam? How large would be the change in the heat of vaporization of water if this amount of energy were not included (internal heat of vaporization)? Ans. (*a*) 167084×10^7 ergs; (*b*) 39.9 cal./g.

35. 500 c.c. of air at 0° and under atmospheric pressure is shut off in a cylinder by a frictionless piston. How much external work is done by the air, considering it as a perfect gas, when it is heated to 100°, and how much heat must be supplied for this purpose? Ans. (*a*) 18.546×10^7 ergs; (*b*) 4.43 cal.

36. Plot Carnot's cycle with entropy as abscissæ and thermodynamic temperature as ordinates. Derive from the diagram the expression for the efficiency of Carnot's cycle.

WAVE MOTION.

By E. Percival Lewis, Ph.D.,

Professor of Physics in the University of California.

339. Characteristics of Wave Motion. The word **wave** recalls the familiar phenomena observed whenever the surface of a body of water is disturbed. Large waves are usually so irregular that it would be difficult to reach any general conclusions regarding the laws of their formation or propagation. If less complex waves be observed, such as those produced by throwing a pebble into a quiet pond or by the gentle disturbance of the water or mercury in a tank, it will be seen that they are alternate ridges and hollows in the surface, which diverge in uniformly expanding circles from the center of disturbance. If small pieces of cork rest on the surface another important characteristic of wave motion may be observed. The particles rise on an approaching wave, ride forward on its crest for a short distance, then fall back into the succeeding hollow, to again move upward and forward on the next crest. They describe orbits in a vertical plane which are evidently circular or elliptical. Since these particles participate in the movement of the water on which they rest, it is plain that the water as a whole does not move continuously forward with the waves, but that each element rotates about its original undisturbed position, to which it returns when the train of waves has passed. Waves are, therefore, the progression of a *shape* or *condition,* not of matter.

340. Water waves illustrate the following fundamental characteristics of all wave motions in material media: (1) *All parts of the medium reached by the disturbance are subject to periodic displacements about their positions of equilibrium.* (2) *The disturbance is propagated at a uniform rate, each displaced particle transferring its motion to its neighbors by pressure or through some mechanical connection.* The moving elements of the medium possess kinetic energy due to their motion and potential energy due

to their displacements. This energy, originally derived from the source of disturbance, is passed on from element to element, so that there is a continuous *flow of energy* with the advancing waves.

341. Types of Waves. The displacements in the case of water waves do not extend far beneath the surface, hence disturbances are propagated in two dimensions only, in superficial waves. There is another familiar type, resembling water waves in general shape, which may be propagated along a linear medium, such as a wire or rope. These may be called linear waves, although the disturbance extends across a finite area, because they are propagated in one direction only. Such waves may be studied by filling a long rubber tube with shot and suspending it from a tall support, holding the lower end taut in the hand. If the tube is struck a sharp blow near the lower end, a distortion resembling a wave crest will travel slowly to the upper end, where it will be immediately reflected with reversed curvature, on account of the elastic reaction at the fixed point. (Fig. 174, *a, b, c.*) It will travel to the lower end and be reflected back and forth several times until its energy is exhausted by friction. This is a **solitary** wave.

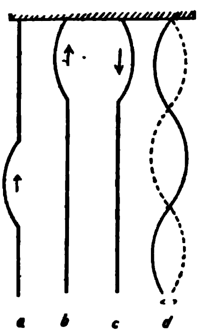


FIG. 174.

If the lower end is rapidly moved back and forth through a small amplitude, by properly timing the displacements a series or **train** of waves of opposite curvatures (" crests and hollows ") will travel upward, crossing a similar train reflected downward. The combined effect of the two trains is to cause the tube to oscillate between the positions shown by the full and the dotted line in Fig. 174, *d.*

In the cases mentioned the oscillations of the medium are in part or altogether at right angles or *transverse* to the direction of propagation, so that the displacements of the boundary of the medium give rise to a definite wave shape. It is possible, however, for the vibrations to take place in the direction of propagation of the wave, as is the case with one component of the displacement in water waves. When the displacements are altogether in the direction of propagation it is evident that the wave can have no shape, as the boundaries of the medium are not displaced, but

there will be periodic changes in density, arising from the fact that different particles are at any instant in different phases of displacement, so that in one region they will be crowded together, while in another they will be separated. This may be illustrated by a row of massive spheres, connected by elastic cords or springs, as shown in Fig. 175, *a*. If the second sphere were immovable, the first alone would oscillate if it is pulled downward and released. If the spheres are all free to move, the transmitted impulse will set all in vibration. On account of the inertia of the spheres and the elasticity of the connections, the displacement of each sphere will lag behind that of its neighbor below, and each vibration will be in a different phase, until we come to the sphere *B*, which begins its first vibration when *A* begins its second vibration. The figure shows the resultant effect when the first sphere has completed one vibration (*b*) and one and a half vibrations (*c*) after it first moved upward through its resting point. It is evident from the figure that the *conditions* of condensation and of rarefaction are propagated with the velocity of the wave. There is no change of shape in the system, but if lines proportional to the displacements are drawn from each resting point, to the right for upward displacements, to the left for downward displacements (that is, if each displacement is rotated through 90° to the right or the left), a smooth curve drawn through the ends of these lines will have the general shape of a transverse wave (*b, c*). We have thus a means of graphically representing *longitudinal* waves in a way clearly coördinating them with *transverse* waves.

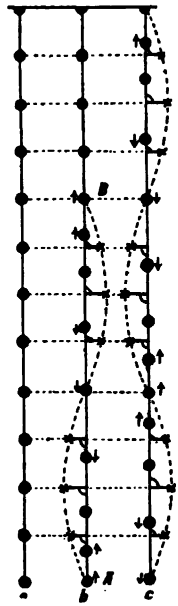


FIG. 175.

If a series of heavy bars are attached horizontally at equal intervals to a suspended wire, and if the lowest bar executes torsional vibrations, waves of angular displacement will travel up the wire. Such *torsional* waves may be represented graphically by erecting ordinates proportional to the angle of torsion at each point on an axis representing the wire.

There are many cases where wave disturbances, such as those

of sound in air, are propagated in three dimensions in a uniform medium. These disturbances will travel equal distances in all directions in equal times, hence the waves will be *spherical*, with the source as a center. A hemispherical wave of this type would be produced in a block of rubber by striking it at a point.

So far we have considered the effect of mechanical disturbances of a medium only. The idea of wave motion may, however, be extended to cases where any physical condition in a medium varies periodically at each point and is propagated with a finite velocity through the medium. A familiar example is found in the "heat waves" which travel into the earth as a result of the periodic heating and cooling of the surface. In the afternoon the surface reaches a maximum temperature. Owing to the slow conduction of the heat, this maximum travels slowly downward, all the while becoming less and less, owing to the fact that each particle passes on only a portion of the energy received by it, not nearly all, as in the case of elastic media. At night the surface reaches a minimum temperature which penetrates into the soil at the same rate as the maximum. The distribution of temperatures in the afternoon and at night are represented by the full and the dotted line in Fig. 176. The abscissa of the point A represents the average temperature. AB is the distance traveled by the heat wave in twenty-four hours. Another example of immaterial waves is found in the electrical waves traveling along conductors or in free space, due to periodic change in the electrical condition at different points. Light waves are believed to be of the same character. (§ 724.)

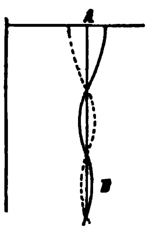


FIG. 176.

In all departments of physics, particularly in Sound, Light, and Electricity, waves play an important part, hence the study of wave motion is of fundamental importance. Since periodic displacements or changes in condition are an essential feature of wave motion, it is necessary to study such phenomena in detail. The only periodic motions which lend themselves readily to simple analysis are those of uniform motion in a circle or the projections of such motions along a line, the latter being called simple harmonic motions. (§ 108 *et seq.*)

342. Simple Harmonic Motion. To summarize the conclusions in the sections above referred to, a body P executing simple harmonic motion with amplitude r and period T moves like the projection of a point C moving with a uniform velocity v in a circle of reference of radius r, with the same period. Hence $v = 2\pi r/T$

$=\omega r$, ω being the angular velocity of the point C. If the time t is counted from the instant at which the radius vector of this point is at an angle e with the axis of reference, the **phase** is $\omega t + e$. (Fig. 177.) To completely describe the motion of the vibrating body we must know the displacement x, the velocity v_x, and the acceleration a_x at any instant t.

$$x = r \sin(\omega t + e)$$

$$v_x = v \cos(\omega t + e) = \omega \sqrt{r^2 - x^2}$$

$$a_x = -\frac{v^2}{r} \sin(\omega t + e) = -\frac{v^2}{r^2} x = -\omega^2 x = -\frac{4\pi^2}{T^2} x$$

As pointed out in § 111, the vibrations of all elastic bodies must

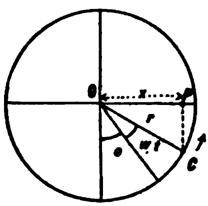


FIG. 177.

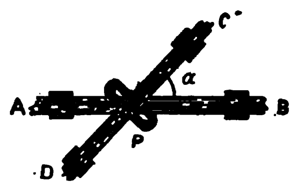


FIG. 178.

be either simple harmonic motions or compounded of such motions (§ 351), since, for small displacements at least, the force of restitution is proportional to the displacement.

343. Resolution of Simple Harmonic Motions. As the motion is a linear displacement, it may be resolved into two or more components like any other displacement (§ 25). If, for example, the piston rod AB (Fig. 178) executes simple harmonic vibrations in a horizontal line (the projected motion of the crank pin on a fly wheel), a pin P attached to it and sliding in a slotted cross bar attached to the rod CD will cause the latter to execute a simple harmonic vibration in the direction of its length, if guides allow it to move only in that direction. If the amplitude of AB is r, the length of the crank arm, that of CD is $r \cos a$.

344. Superposition of Simple Harmonic Motions. In many cases a body may be subjected to several simultaneous simple harmonic displacements in the same or in different directions and of the

same or different periods. Familiar illustrations are found in the vibrations of musical instruments (§ 381 *et seq.*) and whenever different sets of waves are superimposed on or cross each other. If the displacements are entirely independent, it is evident that the resultant effect may be obtained by the geometrical addition of displacements (§ 13). If a light pendulum is suspended from a heavy one, as shown in Fig. 179, and both set in vibration in the same plane, at a given instant the total displacement of the lower bob is $x = x_1 + x_2$; or if the pendulums vibrate in planes at right angles, the resultant displacement is $r = \sqrt{x^2 + y^2}$. In such a case the two systems are not entirely independent, on account of their connections and inertia, and the two displacements will not remain of the simple harmonic type. If a simple pendulum be set in vibration, and later an impulse at right angles to its direction of motion be applied, it will move in a circular or elliptic orbit (conical pendulum), or in a line inclined to its original direction. In studying these effects the most useful cases to consider are those in which the periods of the components are either equal or in some simple ratio to one another.

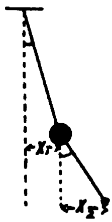


FIG. 179.

345. Composition of Two Simple Harmonic Motions of Same Period and in Same Line. A body at O (Fig. 180) has a simple harmonic motion of period T and amplitude r_1. When the phase of this vibration is e a second simple harmonic vibration of the same period, in the same line, and of amplitude r_2 is imparted to the body. When the phase of the second disturbance becomes ωt that of the first is $\omega t + e$; e is the **phase difference.** To find the resultant displacement and phase, describe circles of reference of radii r_1 and r_2 about O. On these radii, including the angle e, complete the parallelogram, and draw the diagonal $OC = R$.

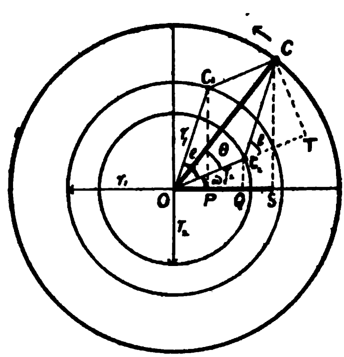


FIG. 180.

Then

$$x_1 = OP = r_1 \cos(\omega t + e);\ x_2 = OQ = r_2 \cos \omega t.$$

$$\overline{OC}^2 = r_1^2 + r_2^2 + 2r_1r_2 \cos e = R^2$$

$$x = x_1 + x_2 = OP + OQ = OQ + QS$$

$$= OS = R \cos(\omega t + \theta)$$

This holds good for any value of ωt. The resultant is, therefore, a simple harmonic motion of the same period as that of the components, of amplitude R, and with a phase $\omega t + \theta$ intermediate between the phases of the components. It is the projected motion of the point C in the resultant circle of reference of radius R. The angle θ may be obtained from the relation

$$\tan \theta = \frac{CT}{OT} = \frac{CT}{OC_2 + C_2T} = \frac{r_1 \sin e}{r_2 + r_1 \cos e}$$

The following special cases are of interest:

$e = 0°$, $R = r_1 + r_2$; $\theta = 0$

$e = 90°$, $R = \sqrt{r_1^2 + r_2^2}$; $\tan \theta = r_1/r_2$

$e = 180°$, $R = r_1 - r_2$; $\tan \theta =$ limiting value of $r_2 \sin e/(r_1 - r_2)$

as e approaches 180°; it will, therefore, be 0° or 180°, according to the sign of $r_1 - r_2$. In the last case, if $r_1 = r_2$, the resultant effect is zero; the body is brought to rest by two equal and opposite displacements.

If the periods are different, e will change uniformly with the time. If the two vibrations start at the same instant, the phase difference at the time t will be $(\omega_1 - \omega_2)t = 2\pi(n_1 - n_2)t$, where n_1 and n_2 are the respective frequencies of vibration. When $(n_1 - n_2)t = 0, 1, 2, 3$, etc., R will be a maximum; when this product equals any odd multiple of one-half, it will be a minimum. The interval between maximum values is $t = 1/(n_1 - n_2)$, and the number of maxima per second is $n_1 - n_2$. This case is illustrated by "beats" in Sound. (§ 379.)

346. Composition of Two Simple Harmonic Motions of Same Period at Right Angles. If the amplitudes of the respective vibrations are r_1 and r_2, construct a rectangle with sides $2r_1$ and $2r_2$, the equilibrium position of the vibrating particle being at the center. Construct two circles of diameters $2r_1$ and $2r_2$, as shown in Fig.

181. The projections on the X and Y axes respectively of points moving uniformly around these circles of reference will give the x and y components of the displacements of the body. If the former is in advance of the latter by the phase angle e, the body will be at P when the y displacement begins. Divide each circle into the same number of equal parts, beginning at C_1 and C_2, and

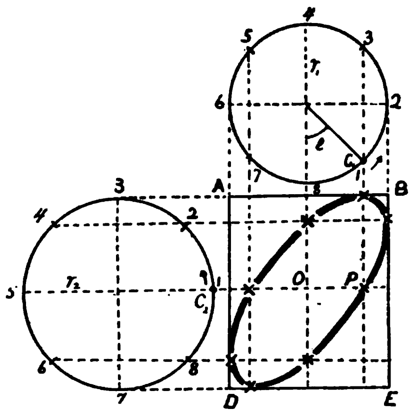


Fig. 181.

number these in regular order. It is evident that the successive positions of the body will be at the intersections of the lines 1–1, 2–2, 3–3, etc., and a smooth curve drawn through these points will give the orbit of the body. In the case illustrated, where $e = 45°$, this path is an ellipse inclined to the axes. If the phase difference is zero, the path is a straight line, the diagonal BD. If $e = 90°$, the path is an ellipse with vertical and horizontal axes, or a circle if $r_1 = r_2$. Orbits corresponding to different values of e are shown in the top row of Fig. 183.

If the periods differ slightly, one vibration will gain on the other in phase, and the orbit will run through the complete cycle of forms shown in the top row of Fig. 183. If n_1, n_2 are the respective frequencies, the cycle will repeat itself whenever one component gains a whole vibration on the other, or $n_1 - n_2$ times a second.

347. Composition of Two Simple Harmonic Motions at Right Angles with Periods in Simple Ratio. Proceed as in the last case, but divide the respective circles of reference into a number of

equal parts proportional to the respective periods, so that the intervals in the two circles will be traversed in equal times. Fig. 182 illustrates the case where $T_1/T_2 = 1:2$, and the angular phase difference is 45° or the time phase difference $\delta = T/8$.

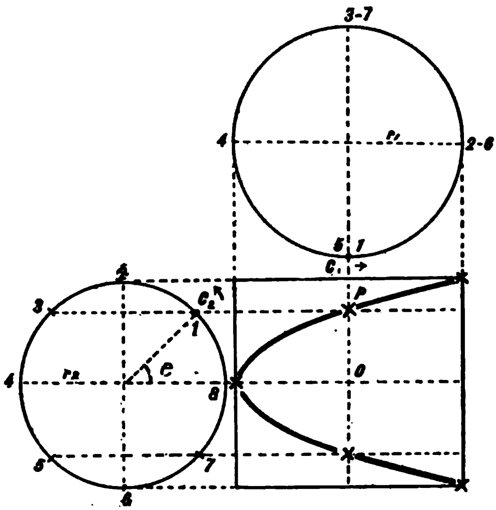


FIG. 182.

Orbits corresponding to other phase differences and to the ratios $T_1/T_2 = 1:3$ and $2:3$ are shown in Fig. 183.

348. Lissajous' Figures. Experimental illustrations of these curves were first obtained by Lissajous. His method was to reflect a beam of light from a mirror attached to one end of a tuning fork to a corresponding mirror on another fork vibrating in a plane at right angles to the first, and thence on a screen. The beam is displaced by both forks, and the spot of light on the screen describes the resultant path. Another method is to use a Y-pendulum, as shown in Fig. 184. If the bob vibrates in the plane of the paper, the effective length is PQ; if it vibrates at right angles to this plane, it is CQ. The periods in the two planes will, therefore, be different and independent. By properly adjusting the lengths PQ and CQ the bob may be made to describe the various Lissajous' figures. If the pendulum has a single support,

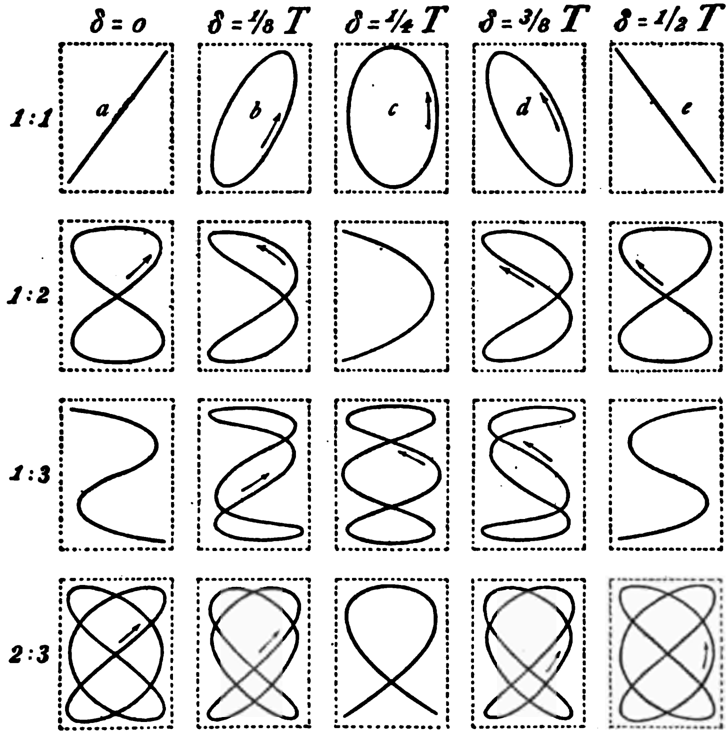


FIG. 183.

$T_1 = T_2$, and the bob will move in an ellipse, circle, or straight line, according to the difference of phase between two impulses given to it at right angles. A rectangular rod fastened at one end will vibrate transversely with a period depending on its thickness. If the diameters parallel to two sides are different, the respective periods of vibration will be inversely as the diameters. If drawn aside diagonally and released, the rod will not continue to vibrate in that direction, but the displacement will be resolved into two components parallel to the diameters. If the ratio of the periods is simple, the end of the rod will describe Lissajous' figures.

349. Waves due to Simple Harmonic Motion. Consider a number of spheres of equal masses attached to each other by elastic

connections, as in Fig. 185. If a transverse simple harmonic vibration is imparted to the first, the impulse will be transmitted to the others in succession. Suppose the phase difference between

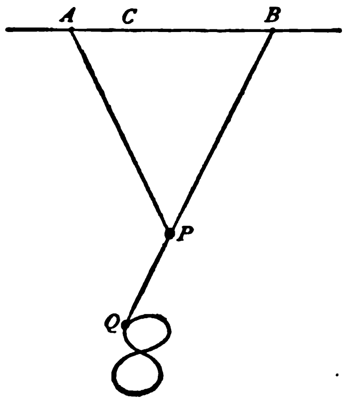


FIG. 184.

the displacements of successive spheres to be one-eighth of a period. When *a* has completed one vibration, *b* has completed seven-eighths of a vibration, etc., while *i* is just beginning to move. The positions of the spheres will be at the projections on the vertical lines 1, 2, 3, etc., of the points 1, 2, 3, etc., of a circle of reference, with radius equal to the amplitude of the wave. If

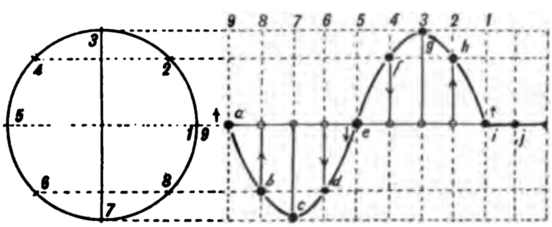


FIG. 185.

a smooth curve be drawn through these positions, it will give the wave form. It is evident that the abscissa of any point on this curve is proportional to the time required for the disturbance to reach that point, or to the phase angle, and its ordinate to the sine of the phase angle of the disturbance at the point. Such

a locus is called a **harmonic curve** or **sine curve**, and gives the shape of a transverse wave when the medium executes simple harmonic vibrations. If the particles in Fig. 175 execute simple harmonic vibrations, the longitudinal wave will be of the same type, and may be represented by a sine curve.

The **period** and the **amplitude** of the wave are the same as those of the simple harmonic motion of any point in the medium. The **wave length** λ *is the distance between any two consecutive points in the same phase of displacement,* for example *a* and *i* (Fig. 185). If V is the velocity of propagation of the wave, $VT = \lambda$, since λ is the distance traversed by the wave during a complete vibration of the "source," sphere *a*. If $n = 1/T$ is the **frequency** of vibration, $V = n\lambda$, the length of the train of waves sent out in one second.

The displacement in a longitudinal wave presents the same aspect if looked at from any direction in a plane at right angles to the direction of propagation. This is not the case with transverse waves, in which the vibrations will be in the line of sight if viewed in the plane of vibration, and at right angles to the line of sight if viewed normally to this plane. Transverse waves have a sort of polarity, therefore, and are said to be *plane polarised.*

Transverse waves may be set up in a cord or longitudinal waves in a spiral spring by fixing one end and attaching the other to a vibrating tuning fork. The amplitude of the waves in such cases may be much greater than that of the fork.

If a beam of light be reflected from a mirror attached to the end of a vibrating fork, and again reflected to a screen from a revolving mirror, the

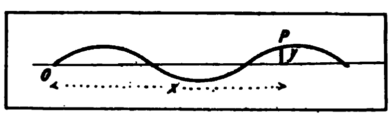


FIG. 186.

harmonic curve will be traced on the screen by the spot of light. Persistence of vision will cause the path to appear continuous.

A permanent record of such curves may be made by causing a bristle attached to the end of a tuning fork to trace its path on the smoked surface of a piece of glass which is moved past the fork at a uniform rate V. The coördinates at the time t of a point P on the sine curve, with respect to the origin O, are evidently (Fig. 186)

$$x = Vt$$

$$y = r \sin \frac{2\pi}{T} t$$

Eliminating t,

$$y = r \sin \frac{2\pi}{T} x = r \sin \frac{2\pi}{\lambda} x$$

This is the equation of a sine curve repeating itself at intervals of $x = \lambda$.

If $y = r \sin 2\pi t/T$ is the harmonic displacement at a given point, the disturbance will reach a point at a distance x in the time $t_1 = x/V$; the disturbance at the point x at the time t will have the phase

$$\frac{2\pi}{T}\left(t - t_1\right) = \frac{2\pi}{T}\left(t - \frac{x}{V}\right)$$

and

$$y = r \sin \frac{2\pi}{T}\left(t - \frac{x}{V}\right) = r \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

This is the equation of wave motion. At a given time t, say $t = 0$, it gives the instantaneous picture of the wave train as a sine curve. At a given point, say $x = 0$, it represents the simple harmonic vibration of the medium at that point.

350. Superposition and Interference of Waves. If two or more trains of waves are superimposed, each will give rise to indepen-

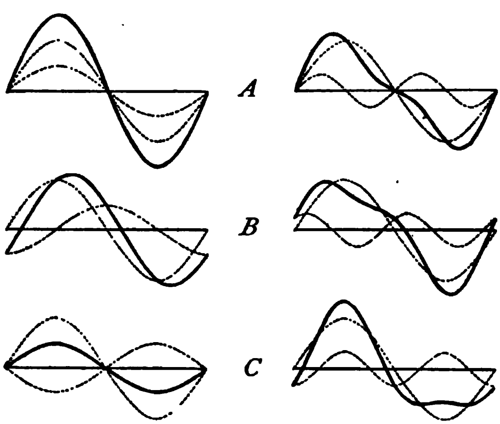


FIG. 187*a*. FIG. 187*b*.

dent displacements of the medium. The resultant effect may be obtained, therefore, by plotting each train of waves on the same axis, with relative displacements corresponding to their phase dif-

ferences, and adding the ordinates. It is convenient to express the phase differences in terms of wave length. If, for example, one wave starts half a period later than another, it should be plotted with its front half a wave length behind that of the first. In Fig. 187*a*, *A*, *B*, *C*, the full line represents the resultant of two waves of the same length and with phase differences of 0, $\lambda/4$, and $\lambda/2$, respectively. In the last case the resultant effect is zero if the amplitudes are equal. The modification of amplitude due to the superposition of waves is called **interference**. It is evident that the length of the resultant wave is the same as that of its components, and that it is a harmonic curve if they are harmonic curves.

351. Complex Waves. Waves of different lengths may be combined in the same manner. If the lengths are in simple ratio to one another, all the resultant waves in a train will be of the same form, but this form will vary with the phase difference, and will not be a sine curve. This is illustrated by Fig. 187*b*, *A*, *B*, *C*, which shows the resultant of two waves of lengths in the ratio 1 : 2, and having different phase relations. Fig. 188 illustrates the case where the lengths are as 1 : 3 and the phase difference zero. These are also examples of interference.

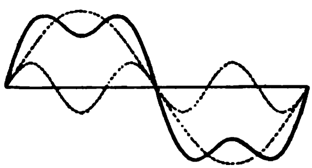


FIG. 188.

If the components have lengths which are not in simple ratio, successive waves will not be of the same shape, as the length of the longest wave will not be a common multiple of the lengths of the component waves.

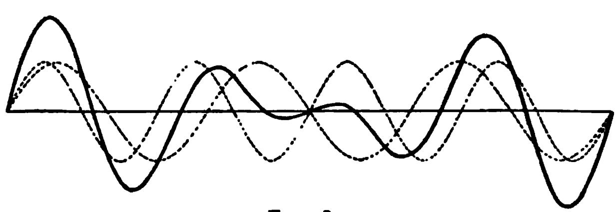


FIG. 189.

If there are only two components, however, with frequencies n_1 and n_2, one wave will gain its own length on the other in $1/(n_1 - n_2)$ second, and the wave train will consist of similar groups repeating themselves $n_1 - n_2$ times a second. The length of each group will be the least common multiple of

the lengths of the components. Fig. 189 shows the effect of superimposing two trains of waves of lengths having the ratio 3 : 4. The graphical representation of "beat" waves in sound would resemble this figure (§ 379). Such forms may be obtained experimentally by the optical method for obtaining sine curves described in § 349, the beam of light being reflected successively from two forks vibrating in the same plane, and giving beats, and from a rotating mirror to a screen.

The displacements of the medium may be the resultant of two displacements at right angles. The example of water waves has already been mentioned. If one end of a cord be attached to the end of a rectangular rod vibrating transversely parallel to each of its diameters, the end of the rod will describe Lissajous' figures (§ 348) and each element of the cord will do the same. If the two diameters of the rod are equal, each element of the cord will move in a circle or ellipse in a plane transverse to its length, but the phases will differ from point to point, so that at a given instant the cord will have the shape of a corkscrew. Such a wave is said to be *circularly* or *elliptically polarized.*

352. Fourier's Theorem. The illustrations given show that various complicated forms may be obtained by the addition of simple harmonic waves of different lengths and phases, and that these waves will be of persistent form if the periods of the components are simple fractions of the periods of the longest component. Fourier proved that any periodic disturbance or wave form of permanent type could be represented as the summation of a number of simple harmonic terms of the form

$$x = r_1 \sin \omega t + r_2 \sin 2\omega t + r_3 \sin 3\omega t + \cdots, \text{etc.},$$

the periods and wave lengths of the components having the ratios 1, 1/2, 1/3, 1/4, etc. Fig. 188 shows that the resultant is approaching a rectangular form, which may be finally attained by adding shorter waves.

The forms of complex waves may be projected by the following device (Fig. 190): A screen with a slit opening at O is placed in front of a horizontal stretched wire AB, which is illuminated by the lens L. An image of the segment opposite the opening may be thrown on a screen S after reflection from a rotating mirror M. If the wire is at rest the image of the illuminated segment will be drawn out in a dark straight line on the screen. If the wire vibrates in a vertical plane the images of the segment in its successive phases as the wave passes O will be laid off end to end on the screen, giving the actual form of the wave passing the opening.

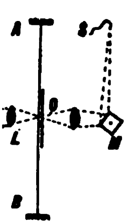


FIG. 190.

353. Velocity of Waves. It might be expected that the velocity of waves in an elastic medium would depend upon the elasticity, which determines the rate at which an impulse is transmitted from one element to another (in a perfectly rigid and incompressible medium the effect would be instantaneous), and the density, which exercises a retarding influence, on account of the inertia of the displaced elements. The derivation of the exact relation between the velocity, the density ρ, and the coefficient of elasticity E is in some cases mathematically difficult, but the general form, at least, is readily obtained from a consideration of the dimensions of the quantities involved (§ 154). If the velocity depends solely on ρ and E, we may write $V = kE^x\rho^y$, where x and y are unknown powers, and k a factor of proportionality. Substituting dimensional expressions for the quantities (remembering that E is force per unit area and ρ is mass per unit volume), we have

$$(V) = \frac{(L)}{(T)} = \left(\frac{ML}{T^2L^2}\right)^x\left(\frac{M}{L^3}\right)^y$$

By inspection we find with respect to T that x must be 1/2. To make M disappear from the right-hand side, y must be $-\frac{1}{2}$. Therefore

$$V = k\sqrt{\frac{E}{\rho}}$$

The exact relation may be easily found in some simple cases. Suppose the front of the disturbance in a longitudinal wave in a medium of unit cross section to be at A (Fig. 191) at one instant and at B a short time t later. The velocity of the wave is, therefore, $V = l/t$, where $l = AB$. An imaginary plane A in the medium is displaced to D, a distance x, by compression. If l is very small, the density of the substance is practically uniform between D and B, and the center of mass of the element is displaced from C to C', a distance $x/2$. The average velocity of the center of mass is $x/2t$ and its final velocity x/t. The final force acting on the element is Ex/l, where E is Young's modulus in the case of a solid, or the pressure p in case of a gas. The average force is half of the above. Equating the work done by this force to the acquired kinetic energy due to the motion of the center of mass, we have

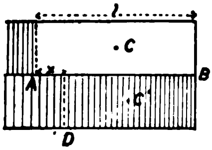


FIG. 191.

$$\frac{Ex}{2l}x = \frac{l}{2}\rho l v^2 = \frac{l}{2}\rho l\frac{x^2}{t^2}\text{; therefore, }\frac{l^2}{t^2} = V^2 = \frac{E}{\rho}\text{ or }\frac{p}{\rho}$$

This applies to longitudinal waves in a wire or rod or to sound waves in a gas.

As another example, consider a transverse wave in a cord under tension. Imagine a bent tube around the cord, as shown in Fig. 192. If the tube moves to the right with the velocity V, the distortion moves with it. The tension T will exert a pressure Tl/R on the lower surface of the tube, if l is the length $CABD$ of the latter, since T is analogous to surface tension (§ 213). If C is the center of curvature of the tube, each element of the cord will be subject to a centrifugal acceleration relative to this point, the entire force acting outward being mlV^2/R, if m is the mass per unit length. At a certain velocity the centrifugal force and the inwardly directed pressure will balance each other, and the curve would preserve its shape and motion if the tube were removed—in other words, it would travel as a wave along the cord. When this is the case,

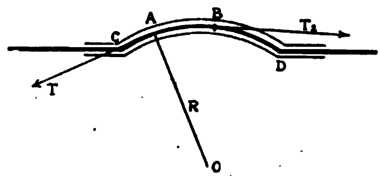


FIG. 192.

$$ml\frac{V^2}{R}=\frac{lT}{R} \qquad \therefore V=\sqrt{\frac{T}{m}}$$

The velocity is independent of R, hence transverse waves of all curvatures and forms travel with the same velocity.

If the diameters of a rectangular rod are different, transverse waves parallel to the different diameters will travel with different velocities, on account of their different periods (§ 348). Therefore any diagonal impulse will resolve itself into two transverse waves which will travel with different velocities.

354. Reflection of Waves. When a transverse wave reaches the fixed end of a cord, the displacement is immediately reversed in direction by the elastic reaction of the fixed end. The wave is, therefore, reflected with reversal of phase of displacement, as shown in Fig. 174*c*. Apparently the incident wave has disappeared through the end, while a wave of opposite displacement has entered, traveling in the opposite direction, and at every instant exactly neutralizing the displacement of the end which would be caused by the incident wave if the end were free. When a continuous train is reflected, the effect is as though a train of indefinite length had been cut in two when a wave reaches A, the fixed point (Fig. 193*a*), and the waves to the right immediately reversed in direction, while the incident waves continue their motion un-

changed; or as though a train of incident waves were traveling through a mirror, while their **inverted** images proceed out of it in the opposite direction.

If one end of the cord is free, when the wave reaches that point, the end, having nothing beyond to restrain it, has an outward displacement twice as great as though the cord were continuous, and it will, therefore, immediately start a wave of the same phase in the reverse direction. After half a period of vibration it will return through the resting point in the opposite direction, and will start a backward wave with phase opposite to that of the incident wave. The case resembles the preceding, except that there is a delay of half a period in the reflection of the wave of opposite phase. It is as though a train has been cut in two at the free end *B* (Fig. 193*b*), and the right-hand section immediately reversed; or at *A* (Fig. 193*a*), and the right-hand section held at rest for

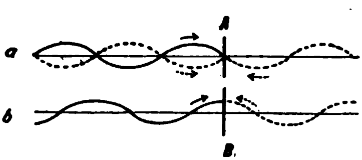


FIG. 193.

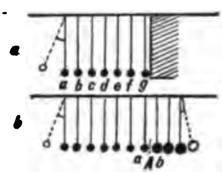


FIG. 194.

half a period before starting in the negative direction, so that the incident wave would then be in the position shown by the dotted line; or in general as if an advancing train were passing through a mirror while its **erect** image emerged from it.

Reflection of longitudinal waves may be illustrated by the conduct of a row of elastic pendulums of the same size as shown in Fig. 194*a*, the last resting against a fixed obstacle. If *a* is drawn aside and released, it will impart an impulse to *b*, this in turn to *c*, etc., and a compression wave will travel to the other end of the row; *g* cannot move, but will be compressed, and through its elastic reaction it will almost immediately start a compression wave in the opposite direction. When this wave reaches the free end, *a* will fly out without restraint, leaving a rarefaction behind it; or, if elastically connected with *b*, it will at once send back a rarefaction wave. In any event, after executing half a vibration it will

swing back through its equilibrium position and reflect a compression wave to the right.

If there are two rows of elastic pendulums of different masses, as in Fig. 194*b*, displacements will be immediately transmitted across *A*, no matter in which direction the wave is moving, but the wave will be also partially reflected. If the wave travels to the right, reflection of *a* at *A* will be immediate; if it travels to the left, the more massive sphere *b* will continue after impact to move to the left, and will return through its resting point, to send a wave back to the right, at the expiration of half its period of vibration.

The cases mentioned illustrate the general principle that the displacements in a medium have a minimum amplitude at a fixed or constrained boundary; a maximum amplitude at a free boundary or one with diminished constraint. Important illustrations of this principle arise in cases where waves pass from a light to a dense medium or *vice versa* (§§ 382, 471).

355. Stationary Waves. Consider a train of waves in a cord moving to the right, while a similar train (reflected or independent) moves to the left. Interference will take place, and the resultant displacement of the medium at a given point and time will be the sum of the individual displacements. Plot the positions of the waves at successive instants (say at intervals of an eighth of a period). If the incident train is represented by a light line, the reflected train by a dotted line, and the resultant by a heavy line (Fig. 195), it will be seen that there are always points of zero displacements *N* (or of minimum displacement if the amplitudes are unequal) at intervals of half a wave length, where the waves always meet in opposite phases. Half way between these points, at *L*, the waves will always meet in the same phase, and the displacement will be a maximum. The former positions are called **nodes**, the latter **loops** or **antinodes.** Between the nodes the medium oscillates back and forth, the direction of the displacements being opposite in adjacent

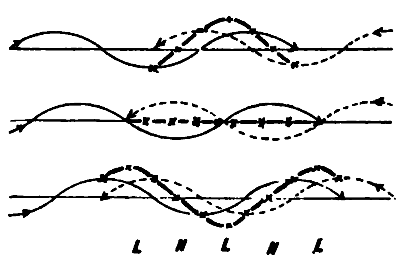


FIG. 195.

segments, so that at any instant the cord has a more or less sinuous shape, except at intervals of half a period, when it passes through the undisturbed straight position (Fig. 174*d*). The same conclusions apply to longitudinal waves. Disturbances of this sort are called **stationary** waves. It is evident that when these arise from the interference of incident and reflected waves there must be a node at a fixed or constrained boundary, a loop at a free or unconstrained boundary.

If one end of a light string is attached to the end of a tuning fork and it is passed over a pulley and stretched by a weight, stationary waves may be obtained by setting the fork in vibration and properly adjusting the length and tension of the cord. If the string is attached to both prongs of the fork (which are always in opposite phases of vibration) by a long Y connection, the waves will not pass beyond the junction because they meet at that point in opposite phases and the resultant effect is zero. If one branch of the Y is pinched by the fingers, the waves in that branch will be destroyed and those in the other branch will now travel to the end.

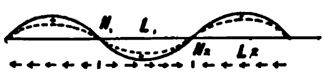


FIG. 196.

Fig. 196 is the graphical representation of stationary waves of longitudinal type. The displacements have just begun to return from maximum elongation, from the full to the dotted line. This indicates that the particles to the left of N_1 and those to the right of N_2 are moving in the negative direction, while those between N_1 and N_2 are moving in the positive direction. Consequently the particles on opposite sides of N_2 are approaching that point, while those on opposite sides of N_1 are receding from it. At N_2 there will be a condensation, at N_1 a rarefaction. After half a period conditions will be reversed. In the neighborhood of L_1, L_2, however, the particles are moving in the same direction with approximately the same velocity, so that their relative positions are only slightly changed. It follows that at the nodes there are the greatest variations of pressure and the least motion; at the loops, the smallest variations of pressure and the greatest motion (§ 382).

356. Waves in a Liquid. Some of the most interesting properties of wave motion may be illustrated by waves on the surface of a liquid, such as water. The initial displacement may arise from differences of level caused by some external force, such as the impact of a pebble, winds, etc. The effect of gravity, of fluid pressure, and of surface tension is to restore the original level, but, on account of their inertia, the particles are displaced beyond their equilibrium positions, just as in the case of vibrations of a

liquid in a U-tube. Horizontal as well as vertical displacements must occur, as in the case of the liquid in the bend of the U-tube. There is, therefore, a longitudinal as well as a transverse component. These displacements are simple harmonic, because the active pressure on an element is proportional to its vertical displacement from the undisturbed surface. We have seen (§ 339) that on a crest the element moves forward, in the hollow backward, in intermediate positions both vertically and horizontally. Fig. 197 shows the positions and directions of rotation of a number of particles originally at rest on the surface in the positions under *a, b, c,* etc., the phase difference between successive displacements being an eighth

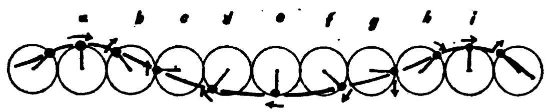


FIG. 197.

of a period. Particle *a* is subject solely to a downward acceleration, particle *e* to an upward acceleration; particles *c* and *g* are subject solely to horizontal accelerations, due to the lateral pressure, as they are in the horizontal plane of equilibrium. We thus find that there is a difference of phase of a quarter period between the vertical and horizontal accelerations, in accordance with the observed fact that the disturbed elements move in elliptic or circular orbits. It is evident that the wave form is not a sine curve.

The expression for the velocity of liquid waves is complicated and cannot be derived here. It is sufficient to say that large waves are maintained by gravity alone, and that the velocity is independent of the density of the liquid, as the force acting is proportional to the weight of the displaced elements, and hence will produce the same acceleration, whatever the density. The velocity increases with the wave length, so that one may frequently see a train of long water waves sweeping through a train of shorter waves and leaving them behind. When the liquid is shallow, the velocity diminishes with the depth. The very small waves are maintained by surface tension alone, so that they are analogous to transverse waves in an elastic membrane. In the case of these waves the velocity increases as the wave length diminishes, and is

also dependent upon the density and the surface tension. Such waves are called **ripples.**

In the case of reflection at a solid vertical boundary, it is easily seen that there must be a node at the boundary for the longitudinal component, but a loop for the vertical component, so that the wave rises against the wall; and that this results in a reversal of the direction of rotation. Stationary waves will be formed, of such a character that at a crest or hollow, the motion will be entirely vertical, at intermediate positions entirely horizontal. The water will alternately pour from two crests into the intervening hollow, until a crest is formed there, and then back in opposite directions into the hollows thus created.

357. Refraction of Waves. Waves move more slowly in shallow than in deep water. Hence if the front AB of an ocean wave moving in the direction of the arrow (Fig. 198) approaches a beach CD, the nearer end, B, of the wave will be retarded more than A, being in shallower water. The wave front will swing around into the successive positions $A'B'$ and $A''B''$, and will finally become parallel to the shore line. This change in direction due to change in velocity is called **refraction.**

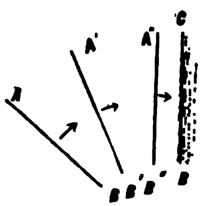


FIG. 198.

358. Propagation and Reflection of Ripples. Experiments with ripple waves may be shown by the following arrangement: A shallow wooden box, about two feet square, is mounted on legs like a table, carefully leveled, and partly filled with water. Light may be projected upward through the bottom from an arc light placed beneath the box, or by reflecting a divergent beam of sunlight upward by an inclined mirror. Ripples on the surface will by their lens effect change the distribution of light on the ceiling so that the motion of each ripple may be followed.

If the middle of the surface is touched with a nail, a circular ripple will diverge from that point. If the surface were larger, this wave would at a later time occupy the position of the circle (Fig. 199), but it will be in part reflected from the four sides. The reflected segments are exactly like the missing segments of the outgoing wave, reversed in direction. These reflected waves have centers at C_1, C_2, C_3 and C_4, the "images" of the source C, which are evidently at the same distance from the walls as the source itself, since C and the other centers of curvature are symmetrically

situated with respect to the walls. These reflected waves will cross each other and be subject to repeated reflections ("multiple reflection"), their curvature all the while decreasing, until we have a rectangular system of straight ripples.

If a circular wave strikes a bent sheet of metal of the same

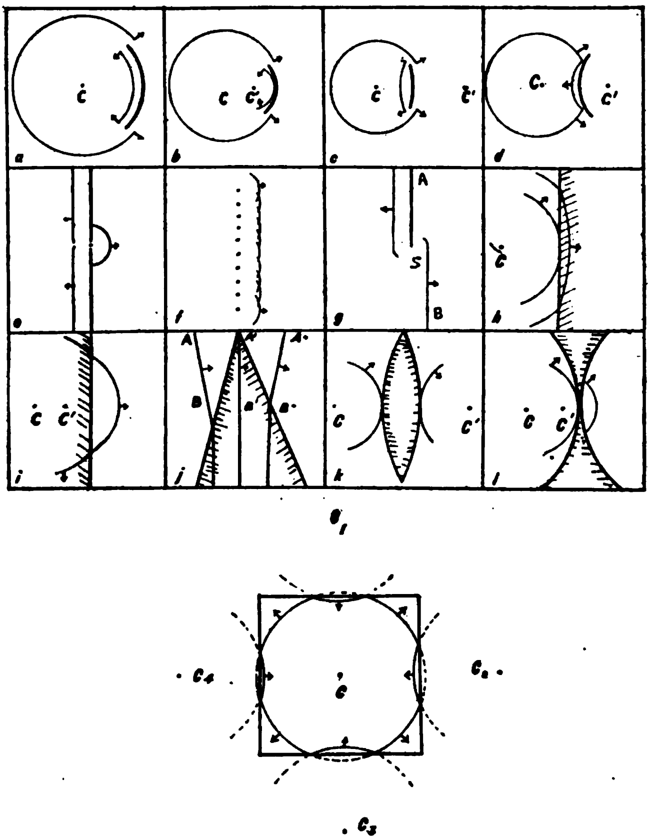


FIG. 199.

curvature as the wave, the latter will be reflected without change of curvature, converge to its starting point, and diverge from it on the opposite side (Fig. 199*a*). If the strip has a greater curvature than the wave, the edges of the latter will be first reflected,

so that its curvature is increased (b). It will converge to C', a "real image" or "focus." If the strip is concave with less curvature than the wave (c) the latter may be made divergent, with a "virtual" image at C'. If the strip is convex toward the wave, the reflected wave will always diverge from a virtual center behind the mirror (d).

If the surface is touched with a long straight strip of metal a straight ripple will be produced. If this strikes a screen with a small slit in it (e) the disturbance will pass through this hole and set up a semicircular wave on the other side. The remainder of the wave will be reflected as a straight line.

If a number of nails are driven at equal distances through a strip of wood and dipped into the water, circular waves will diverge from the points of contact. At a little distance these wavelets will blend into a straight ripple corresponding to their common tangent (f). At other points the ripples cross each other in all phases, and their effect will vanish because of interference. We may, therefore, consider that a linear wave front is due either to a continuous linear disturbance or to a number of neighboring point disturbances, each sending out circular waves. In (e) for example only the point in the opening is effective for transmission. The latter conception is often useful. (§ 420.)

If a screen S projects part way across the tank (g), the portion AS of an incident wave will be reflected; the remainder SB of the wave will pass the screen. It will be noted that the end of transmitted wave front will bend into the shadow of the screen, and the end of the reflected wave will bend into the region formerly occupied exclusively by the other half of the wave. S is apparently a center of disturbance for both these waves. This effect is called **diffraction.** By noting the resemblance of the ends of the waves in this case to those in the preceding case (f) the explanation will be made clear.

359. Refraction of Ripples. Advantage may be taken of the fact that the velocity of water waves diminishes with the depth to illustrate refraction. On the bottom of the tank lay a piece of thick glass, so that the water over it is about one-fourth as deep as elsewhere. A linear ripple is started by touching the surface with a strip of metal. On reaching the edge of the glass plate

the end B is retarded and the wave will swing into the position $A'B'$, as in Fig. 198.

If the incident wave is circular, the middle will be more retarded than the edges if the wave comes from the deeper water, and the curvature of the wave will be diminished (h). If the wave travels from the shallow region, the contrary will be the case (i). The centers of curvature or "images" of the source will be at C' (outside the tank in h).

If a prismatic sheet of glass is laid on the bottom (j) a linear wave front AB will be rotated both in approaching and leaving, and the final direction will be $A''B''$. If pieces of glass with convex or with concave edges, like sections of lenses, and laid on the bottom, the center of a passing circular wave will be more retarded than the edges in the first case (k), less retarded in the second (l), resulting in changes of curvature. The "images" of the source will be at C'.

360. Interference of Ripples. If two nails simultaneously touch the water at different points two circular waves will be set up, which will cross and interfere with each other. They pass so quickly, however, that it is difficult to observe them. Better results will be secured if a continuous series of waves can be produced, and still better results if there is a system of stationary waves. A very satisfactory method of securing this result is to put mercury in a circular glass dish at least four inches in diameter, and maintain periodic disturbances at the center by a glass fiber attached to the vibrating prong of a tuning fork. Continuous trains of circular ripples will diverge from the center, while reflected circular ripples will converge toward that point. The result will be a system of circular stationary waves, as illustrated in Fig. 200. They may be projected on a screen by reflected light, and made more distinct by using a lens.

If two glass fibers are attached to the fork near each other, two trains of waves will be maintained, and each will form its own system of stationary waves. At all points on the surface where the outgoing waves meet each other in the same phase (that is, where the difference of the respective distances to the two sources is zero or any whole number of wave lengths), the waves will reinforce each other. In regions where they meet in opposite phases

(the differences of path being some odd multiple of a half wave length), they will destructively interfere with each other. Along certain lines, therefore, there will be no disturbance by either outgoing or reflected ripples (Fig. 231). Between these lines segments of the stationary waves appear, as shown in Fig. 200.

361. Energy and Intensity of Waves. The energy of a vibrating body is proportional to the square of its amplitude (§ 61). Each

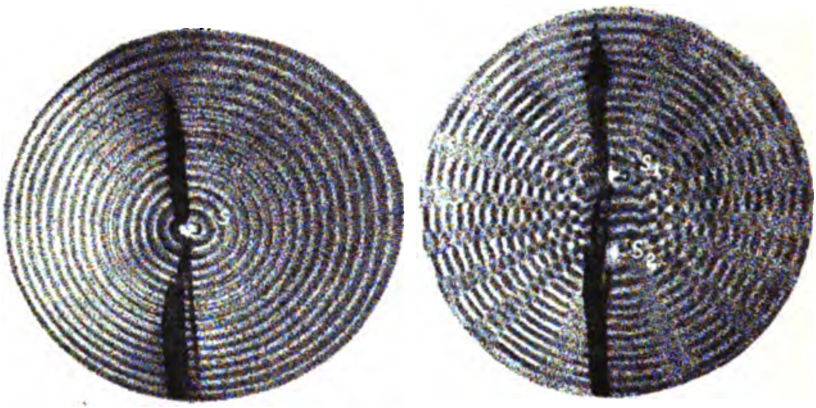


FIG. 200.

vibrating element of mass in a medium traversed by waves will, therefore, possess energy proportional to the square of its amplitude, and this energy will flow forward with the advancing waves. The **intensity** of waves in a given region is defined as being proportional to the amount of energy passing per second through unit area at right angles to the direction of propagation; hence the intensity is proportional jointly to the amplitude and the velocity of the waves. When they travel in a viscous medium, such as molasses or lead, they rapidly decay in amplitude and disappear, owing to the absorption of energy by internal friction. This effect is known as **damping**. Fig. 176 represents the form of a damped train of waves. If there is no such loss the same quantity of energy will persist in a given wave, no matter how far it travels, or how the dimensions and form of the wave front may change. If such waves travel in a wire or any other channel of constant cross section the intensity will be independent of the distance from

the source, as the wave front will remain of constant area. This is illustrated by the transmission of sound waves through a speaking tube or of light waves in a parallel beam. In the case of circular waves on a surface, a constant amount of energy will remain in a wave of circumference which increases directly as the distance from the source; hence the intensity must vary inversely as the distance, and the amplitude inversely as the square root of the distance. In the case of spherical waves, the energy will remain constant within a spherical shell of the thickness of one wave length and with surface increasing as the square of the distance. If E is the energy emitted from the source per second, and if r_1 and r_2 are the radii of the wave at different distances, and I_1 and I_2 the corresponding intensities,

$$E = 4\pi r_1^2 I_1 = 4\pi r_2^2 I_2 \quad \therefore \frac{I_1}{I_2} = \frac{r_2^2}{r_1^2}$$

Hence the intensity varies inversely as the square of the distance and the amplitude inversely as the distance.

REFERENCES.

FLEMING'S *Waves and Ripples in Water, Air, and Ether* is an excellent popular description of all kinds of waves.
EDSER'S *Light*, chapters on Wave Motion.
DANIELL'S *Principles of Physics*, chapters on Wave Motion.
WOOD'S *Physical Optics*, Ch. 3 and 4, gives an interesting account of the photography of sound waves.

PROBLEMS.

Simple Harmonic Motion.

1. A mass of 196 grams is suspended by a rubber band of such elasticity that an additional weight of 5 grams will stretch it 1 cm. It is extended 1 cm. and released. Find the period, and the displacement, velocity, and acceleration 9 seconds after it passes upward through its resting point.

Ans. $T = 1.256$ sec.
$x = \sin 59°.6 = 0.862$ upward.
$v = 2.45$ cm./sec. upward.
$a = 21.55$ cm./sec.^{2} downward.

2. Water or mercury in a U-tube is disturbed. Show that the liquid executes a simple harmonic motion of period $T = 2\pi\sqrt{l/g}$, where l is the length of liquid from surface to surface around the bend.

3. Compound two simple harmonic motions of same period and in same plane with amplitudes 3 and 2 and with phase difference of one-sixth of a period. Ans. $R = 4.36$.

4. Compound two simple harmonic motions at right angles with periods in the ratio 3 : 5 and with phase difference zero.

Waves.

5. Compound three trains of waves of lengths in the ratios 1, $\frac{1}{3}$, and $\frac{1}{5}$ and of amplitudes 3, 2, and 1, starting in the same phase.

6. Compound two trains of waves of lengths 5 and 4 and of equal amplitudes.

7. A copper wire ($\rho = 8.8$) 1 square mm. in cross section is subject to a tension of 88,000 dynes. With what velocity will a transverse wave travel in it? Ans. 1000 cm./sec.

8. With what velocity will a longitudinal wave travel in the same wire? Ans. 350,000 cm./sec.

SOUND.

By William Hallock, Ph.D.

Professor of Physics in the Columbia University, New York, N. Y.

NATURE AND PROPAGATION OF SOUND.

362. Nature of Sound. The word sound is often used to designate the sensation peculiar to the ear, but, as used in Physics, the word denotes the air disturbances which reach the ear and cause the sensation. When the sources of sound are considered, it is noticed that they perform elastic vibrations and, therefore, give rise to elastic waves or impulses, which are transmitted in all directions through the air or some other medium. A violin string, the air in an organ pipe, and the prongs of a tuning fork are sources of such elastic vibrations. All elastic bodies transmit sound, and matter in some form is necessary for its existence and transmission. An electric bell operating in a vacuum, under the bell jar of an air pump, is not heard, because there is no elastic medium in which to form and transmit the impulses.

It was early recognized that sound is somewhat analogous in its method of propagation to waves on the surface of water, and consequently the term wave was applied to the elastic impulse which constitutes sound. It is, however, evident that sound waves must differ from water waves in one important respect. Waves on the surface of water consist of partly transverse and partly longitudinal vibrations (§ 356). The elastic waves transmitted by air which we call sound cannot consist even in part of transverse vibrations since a gas has a zero shear-modulus (§ 182), that is, it develops no force of an elastic nature that would resist a transverse strain or shear. Sound must consequently consist of waves of **longitudinal vibrations,** and along the path of the waves the medium is alternately condensed and rarefied (§ 341). This conclusion is fully confirmed by observation and experiment.

363. Velocity of Propagation in Air. Although it is a matter

of every-day observation that sound requires an appreciable time to travel from the source to the ear, when they are separated by even a few hundred yards, it was not until the seventeenth century that experimental determinations of this velocity were made. Near Amsterdam the Dutch scientists, Moll and Van Beek, in 1823 stationed two cannons on two hills, each in plain view of observers on both. The cannons were fired simultaneously and the observers noted the time which elapsed between the seeing of the flash and the hearing of the report. In this way a long series of determinations was made, and the observation of the velocities in both directions tended largely to eliminate the effect of the wind. The distance between the stations was 17669 meters and the average elapsed time was 52.07 seconds, giving a velocity of 339.3 m./sec. or 1112.5 ft./sec.

Similar observations were made in 1822 between Villejuif and Monthlery, south of Paris, with Prony, Arago and Matthieu at one station and Humbolt, Gay-Lussac and Bouvard at the other. The distance was about 18.6 kilometers and the time of transit 54.6 seconds, giving a velocity of 340.7 m./sec. or 1116.8 ft./sec.

In these and other determinations it has been found that *the velocity of sound in any medium is practically independent of the pitch of the sound* but depends on the temperature and relative humidity of the air. We shall consider the explanation of these facts later (§ 365).

364. Velocity of Propagation in Water. Colladon and Sturm in 1827 stationed two vessels in Lake Geneva at a considerable distance from each other and so suspended a bell from each vessel that it hung below the surface of the water, and made a series of observations entirely analogous to that of Moll and Van Beek. By this means the velocity of sound in fresh water was experimentally determined, and found to be 1435 meters per sec. Submerged bells are now used very extensively by ships and off shore light vessels for signalling at sea, after the manner of Colladon and Sturm. The transmission of signals through the water is found to be much more reliable than that through the air, the latter being seriously affected by weather conditions. Although special telephone receivers have been devised to immerse in the ocean and pick up the signal, generally the vessel acts as a sounding board and reports the sound satisfactorily. The velocity of sound in solids

has never been determined by this direct method, but means are available for the indirect experimental determination of the velocity of sounds in solids (see § 385).

365. Velocity of Propagation in Gases. It is possible to calculate from theoretical considerations the velocity of propagation, as soon as the elasticity and density of the medium are known. A formula was derived by Newton for the velocity, v, of propagation of a sound wave in a gas in terms of its modulus of elasticity E (§ 223) and density ρ, namely

$$v = \sqrt{\frac{E}{\rho}}$$

This formula was, however, found by Newton himself to be inadequate, since it gave a result about 20 per cent. smaller than that observed experimentally. The explanation of the discrepancy was first given by Laplace, who pointed out that Newton assumed the temperatures in the condensations and rarefactions to remain constant, and hence used the modulus of elasticity of a gas at constant temperature (§ 223). In fact, the temperature is temporarily elevated in a condensation and lowered in a rarefaction, and the proper modulus of elasticity to use is that found by compressing a gas without allowing the heat produced to escape. This is the adiabatic modulus (§ 322) which is γ times greater than the modulus at constant temperature, γ being the ratio of the specific heat of the gas at constant pressure to that at constant volume. The introduction of this factor reconciles theory and experiment, and we shall therefore use the formula derived by Laplace,

$$v = \sqrt{\gamma \frac{E}{\rho}}$$

This formula for gases immediately enables us to predict the conditions which affect the velocity of propagation in gases. For example, the modulus of elasticity of a gas (at constant temperature) being equal to its pressure (§ 223) and the pressure being by Boyle's Law (§ 221) proportional to the density that is E and ρ are proportional to each other, the velocity of sound in a compressed gas is the same as in a rarified gas at the same temperature. The velocity of propagation in higher altitudes where the air is thinner is (at the same temperature) the same as at the

sea level where the air is more dense. Variations in barometric pressure alone do not affect the velocity of propagation.

On the other hand, raising the temperature of a gas and leaving its pressure constant decreases its density, while leaving its elasticity unchanged. Consequently sound will travel more rapidly in warm air than in cold. If ρ_0 be the density of the gas at 0° C. and ρ its density at temperature t° C., since density varies inversely as volume, $\rho_0 = \rho\ (1 + .003665t)$ (see § 256). Hence the velocity at t° C. is

$$v = \sqrt{\frac{\gamma p}{\rho_0}(1 + .003665t)}$$

where p is the pressure of the gas (in dynes per cm²), ρ_0 its density at 0° C. and γ the ratio of the specific heats of the gas.

The general formula also tells us that for different gases, as oxygen, hydrogen, nitrogen, etc., for which γ is the same, the velocity of propagation under the same pressures is inversely proportional to the square root of their density. Thus, for example, the velocity in hydrogen is four times that in oxygen, and nearly four times that in air, the densities being as 1 : 16 : 14.4. Another condition which affects the velocity of propagation in the atmosphere is the amount of water vapor present. Since the density of water vapor is only about two-thirds that of air, its presence tends to increase the velocity of propagation. If the hygrometric state is accurately known its effect upon the velocity can be calculated. The fact that the velocity of propagation is practically independent of the pitch of the sound, at least within wide limits, is in accord with the above formula, since it contains no term depending on the pitch.

Velocity of Sound.

Air	333	meters per second
Hydrogen	1268	" "
Carbonic acid gas	261	" "
Fresh water	1435	" "
Sea water	1454	" "
Mercury	1484	" "
Steel	4975	" "
Lead	1420	" "
Glass	4860	" "

Pine wood	3300	meters per second	
Walnut wood	4800	"	"
India rubber	5000	"	"

366. Reflection. The reflection of sound takes place wherever the sound waves strike upon a surface that is large in comparison with the length of the wave. The phenomenon is familiar in the case of the echo, where the sound is reflected from the side of a building, the cliff of a mountain, the trees at the edge of a forest or the like. The laws of reflection of sound are not easily demonstrable experimentally on account of the diffraction (§ 368) which practically always takes place. Nevertheless, it has been shown, particularly with very short waves, that the angle of incidence is equal to the angle of reflection. This can be qualitatively observed by standing perpendicularly in front of a building and trying the echo in that position, and then moving off to a line oblique to the surface, when there will be comparatively little echo returned to the observer. The sound produced by the discharge of lightning reflected back and forth between the clouds, and between the clouds and the earth constitute the familiar phenomenon of thunder. The reflection of sound is often made use of to estimate the distance of some wall or cliff, because the sound requires a certain time to travel from the observer to the cliff and back, and if this time is observed, a very fair approximation of the distance can be made. Inasmuch as the reflection of an elastic impulse takes place whenever the impulse encounters a change in the elasticity or density conditions of the medium, it follows that sound waves are reflected, not only when they strike upon a hard object when traveling in the air, but also when, traveling in a solid, they reach the end of that medium.

The repeated reflections of sounds from the walls of an auditorium, theater or church may produce a "reverberation" which interferes with distinct hearing, though it may help to deliver sound of greater intensity to the more remote auditors. More or less spherical or ellipsoidal ceilings are generally responsible for "whispering galleries" and the like. Sabine has developed some general principles which can be applied to the design of a successful auditorium. To reduce reverberation the walls should be covered with materials which will absorb a large percentage of the sound energy falling upon them. The amount and duration of the reverberation depend only on the volume of the auditorium, the nature of the walls and the pitch of the note. Sabine found that wooden or plaster walls reflect

90 to 95 per cent. of a sound, and various cloths, hung loose, reflect 75 to 85 per cent. In the latter case the waves pass through the cloth and are reflected from the wall, being only slightly absorbed by the cloth. Tufts has shown that porous fixed partitions transmit sounds in the same proportion as they do air currents, and that loose hanging draperies are moved back and forth with the waves and really absorb very little energy.

367. Refraction. Much that has been said with reference to the uncertainties in the quantitative study of reflection applies to the question of refraction of sound. Nevertheless, careful experiments have shown that the refraction of sound follows laws entirely similar to those governing the refraction of light and other wave motions (§ 405). Comparatively few cases of ordinary observation illustrate the refraction of sound; but it has been observed, particularly at sea, and was investigated by Joseph Henry for the Light House Board in 1860. It was shown in this series of experiments that the sound of a steam siren often passed over the top of the nearer vessel and was heard distinctly by the vessel farther from the source than the one which reported it as silent. Apparently the varying temperature of the air immediately over the water produced a curved path in the propagation of the sound which went completely over one vessel. These curved paths in sound transmission are entirely analogous to those in light producing "Mirage" (see § 467).

Wind may also cause a change in the direction of propagation of sound waves as shown by Fig. 201, where the arrows indicate the direction of the wind and the straight lines represent the wave fronts. In *A* the wind is opposite to the direction in which the sound is traveling and, since the

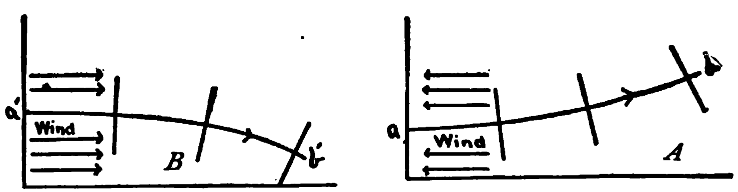


FIG. 201.

wind-velocity is less close to the surface of the earth, the lower parts of the waves travel faster than the higher parts. Hence the sound is deflected upward and may pass over the head of an observer. In *B* where the wind is in the direction of the propagation of the sound the waves are deflected downward.

368. Diffraction. This is the term given to the phenomenon observed when waves do not confine their path of propagation to straight lines but bend around obstacles. In fact, the presence of diffraction in sound usually conceals the phenomena of reflection and refraction. Whereas we are unable to see around a corner, we are able to hear around one with almost undiminished ease. The prominence of diffraction in sound observations is due to the fact that the waves of sound, being usually several feet in length, are comparable in size with the objects which they encounter. A very large obstacle such as a hill has sometimes been observed to cast a comparatively sharp shadow of a loud sound (such as an explosion) when the observer was at a considerable distance from the obstacle. In fact, light waves and sound waves show very similar amounts of diffraction, when the sizes of the obstacles and the distances of observation are in proportion to the lengths of the respective waves.

LOUDNESS, PITCH AND QUALITY.

369. Musical Sounds and Noises. Under the name of noises are classed all those sounds which are too brief or too irregular to have any assignable pitch. Sounds that continue uniform for an appreciable time and have an assignable pitch are called "musical."

Musical sounds differ from one another in three essential factors, namely, **loudness** or amount of sound; **pitch**, that is, highness or lowness of tone; and **quality**, or character of sound, *e. g.*, the sound of a piano, of a violin and of the human voice differ in quality, although they may be identical in loudness and pitch. These three factors completely define any sound and we shall discuss them individually. Three factors also completely define any wave of a given type, and in the case of sound waves it will be found that the **length** corresponds to the *pitch*, the **amplitude** or extent of vibration to *loudness*, and the **shape** of the wave to the *quality* of the sound.

370. Loudness. The least interesting and important of the three factors determining a sound is loudness, because mere loudness is generally a question of nearness to the source, or the amount of energy which is converted into sound waves. Hence it is evident that the comparative loudness of two sounds that are alike in other respects depends on the intensity of the wave-motion,

that is, on the square of the amplitude (§ 361). Theoretically the intensity of sound waves decreases inversely proportional to the square of the distance from the source, as in the case of all spherical waves (§ 361). Nevertheless, owing to the internal friction of the air and other causes, sound vibrations are gradually "damped" and the intensity of the waves at a distance from the source, experimentally determined, falls considerably below the theoretical value. Owing to the structural limitations of the ear the *loudness* of the *sensation* is not always proportional to the energy in the wave motion. The maximum loudness for a given consumption of energy is observed when the pitch is about the middle of the piano scale.

It is possible to increase the effective intensity of sound waves by collecting them at a given point. For example, a hollow cone will intensify by concentrating at its apex the waves entering its open end. The ear trumpet is a familiar example of such a device. Conversely a sound originated in the apex of such a cone is directed essentially along the axis and the effect at a distant point near the axis is greatly increased. This appliance is popularly known as a megaphone.

371. Pitch. The methods available for the measurement of the rate of vibration of a given sound may be divided into two classes. In the first the number of impulses per second is measured directly, and in the second the wave length is the object of direct observation and the frequency is calculated.

The siren (Fig. 202) is adapted to quantitative as well as qualitative tests of pitch. In its usual simple form it consists of a windbox *B*, a revolving disk *D*, and a counting device W. E. The disk *D* is perforated with a number of holes, *e. g.*, 16 equally distributed in a circle around the axis *S*. *D* runs very close to the top of the box *B* in which is also a ring of holes exactly opposite to those in *D*. The holes in the two plates are inclined in opposite directions, as shown at *C*, which is an enlarged section perpendicular to the radius, and passing through any two holes. The stream of air issuing from the lower hole impinges against the side of the hole in *D* and, issuing therefrom, reacts against *D*, both actions tending to cause *D* to rotate about its axis in the direction of the arrow *F*. When *D* moves a little the issuing air is cut off until the holes again come over each other, when another impulse is

given to D and another puff of air allowed to escape. Thus the air furnishes the motive power as well as the puff; the higher the pressure of air the faster D revolves and the higher the pitch rises. If there are 16 holes in D, evidently they will coincide with the

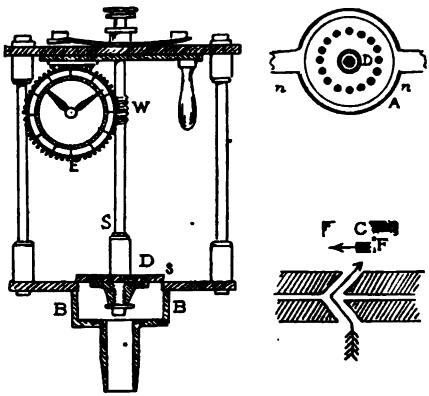


FIG. 202.

holes in B 16 times during a single revolution, thus allowing 16 consecutive puffs of air to escape during one revolution of D.

The siren is so adjusted as to emit a sound identical in pitch with that to be determined, and then the counting device W. E. is thrown into action. The number of revolutions of the plate D in a period of several seconds is observed. Evidently the number of revolutions of D per second multiplied by the number of holes in D, will give the number of air puffs per second, *i. e.*, the frequency of the vibrations imparted to the air.

If a light stilus *r* Fig. 203, be attached to the prong of a tuning fork, *B*, and lamp-blacked paper be made to pass under the stilus at a uniform known velocity, a wave line will be traced on the paper and the number of waves recorded per second will be the rate of the fork. The fork and stilus are connected to one pole of a battery through *P* (Fig. 204), and the drum and frame to the other pole through *N*. A clock is introduced into the circuit so as to start a current through a spark coil every second, or half second, and a spark is thus caused to pass

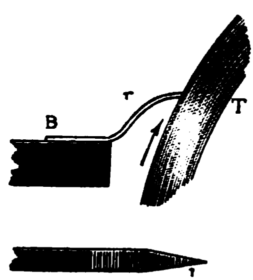


FIG. 203.

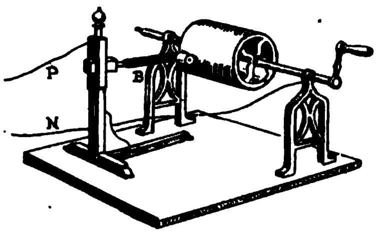


FIG. 204.

from the stilus to the drum and make a mark which automatically records the speed of the drum. This method may be advantageously used to compare the rates of two forks, recording on the same paper. In this case uniform known velocity is not essential since the ratio of the numbers of the waves traced in any given interval of time is the ratio of the rates of the forks.

Determinations made by the above means show that the pitch of a note depends only on the frequency of the source. Notes that are described as higher in pitch have a greater frequency than those described as lower.

If a fork emit impulses at the rate of n per second it will give n waves per second, and if the wave-length is λ, $v = n\lambda$ (§ 349). Hence if we can measure λ by some independent method (§ 382) the value of n can be deduced. The above formula, taken in connection with the statement (§ 363) that the velocity of sound in any one medium is independent of the pitch, forms the proof of the statement made in § 369 that the pitch of a musical sound depends on the wave-length; for evidently, v being a constant, n varies inversely as λ.

372. Doppler Effect. If the ear A is moving with a velocity v' toward a source of sound at B then more waves will reach A in each second than if A were stationary; that is, the pitch on approaching a source will appear higher. The additional waves received by A are those that occupy the distance v' that A moves in a second towards B and these are v'/λ in number or $v'n/v$. Hence the pitch rises from n to $n(1 + v'/v)$. Similarly if A is receding from B, fewer waves reach the ear and the pitch is lowered to $n(1 - v'/v)$. Analogous results follow if A is stationary and B is approaching or receding. This phenomenon is called after its discoverer, the "Doppler effect." It can be readily observed that the pitch of the whistle or bell of a passing locomotive or gong of a trolley car drops as the source of sound passes the observer and changes from approaching to receding. A similar phenomenon is observed in light waves from a source moving with relation to the observer.

373. Limits of Audibility. Whereas impulses in the air following each other less frequently than 10 to 15 per second in general produce the sensation of distinct puffs, when these impulses become more frequent they blend to produce a steady note, and as they become more and more rapid the pitch of the note becomes higher and higher until a point is reached where the ear no longer perceives the sound. In other words, the frequency may become so high that the sound is inaudible. This upper limit is generally in the neighborhood of 30,000 per second; nevertheless, both of these limits vary greatly with different individuals, as does also their ability to estimate pitch, and musical interval, throughout the whole range of sounds. The analogy between pitch in sound and color in light is very close. The nature of the sensation in both cases is a question of the length of the corresponding wave. Just as the rainbow extends farther into the violet or the red for some observers than for others, so some hear higher pitches and others lower than the normal. And, just as we have color blind people unable to distinguish between the different colors of the spectrum and their relations to each other, so we have tone deaf people who are unable to estimate musical interval or relative pitch. Most insect noises are very high pitched, and it is probable that some insects produce sounds that are above the upper limit of audibility of man's ear. A sound of moderate pitch will also be inaudible if the amplitude of vibration is too small. Lord Rayleigh has found that the faintest audible sound has an amplitude of vibration of the air particles of about 10^{-7} cm. In an extremely loud sound such as a steam whistle heard close at hand the amplitude of vibration is probably less than one mm.

374. Quality. The quality of a musical sound is the property that distinguishes two musical sounds of the same pitch and intensity which come from different sources. The qualities of the tone from a violin, piano, flute, etc., differ greatly though they may all play the notes at the same pitch. Two human voices differ in quality.

Since, as we have already seen, the loudness of a sound depends on the amplitude of the wave while the pitch depends on the wave-length, it follows that the third characteristic of a sound, its quality, must depend on the third property of waves, namely,

wave-form. Now we have already seen that the form of a wave (that is not simple harmonic) depends on the waves of higher frequency that are combined with the fundamental. Hence *quality depends upon the pitch and relative intensities of the waves of higher frequency which accompany the fundamental.* When the wave form is simple harmonic it gives rise to the sensation called a *pure tone,* which the most musical ear cannot analyze into constituents (a law discovered by G. S. Ohm). When the wave is complex harmonic (§ 351) a well trained ear can resolve it into constituent pure tones, thus accomplishing in a practical way what Fourier's theorem (§ 352) enables us to do mathematically.

In one group of cases the ear does not distinguish a difference of quality where there is a difference of wave form. Since the ear is capable of analyzing a complex musical sound into its components, so that the hearer can say that certain pure tones are present, it is evident the difference in two waves such as those in Fig. 187*a* or Fig. 187*b*, makes no difference in the quality of the sound, because they contain the same elements but in different phase relation.

375. Musical Interval. Two musical sounds differing in pitch are said to be separated by a certain *interval.* The human ear recognizes certain intervals as pleasing or *harmonious;* others as *inharmonious* or *discordant.* Since the very early days of Greece, the interval of the *octave* has been recognized as the simplest and most pleasing. The *fifth* and *major third* are also pleasing intervals and most persons after a little practice are able to detect these intervals without difficulty, that is, given any tone, they can immediately determine the third or fifth above it by the ear. When the rates of vibration of two tones separated by the interval of an octave are determined experimentally, it is found that the higher tone has just *twice* as many vibrations per second as the lower one. No matter what the actual number in either case, the *ratio* of the two rates is always 2 : 1. Similarly the fifth gives a ratio of the component rates of 3 : 2; the major third of 5 : 4, etc. Thus it is evident that musical interval is a question of the *ratio* of the rates of the components and not of the absolute value of the rates. In general the simpler the ratios, the more agreeable the interval.

376. Musical Scales. If some tone be taken as a starting point

or fundamental, and various ratios or intervals applied to it, a series of tones may be interpolated between that fundamental tone and its octave. The interval between a tone and its octave thus divided into more or less equal intervals is called a *scale*. If the usual musical nomenclature is used and C taken as the fundamental, the interval of the fifth gives G; that of the major third gives E; and so on. The following table gives the ratio to the fundamental, as well as the ratio to next neighbor, of each tone when the interval of the octave is subdivided in the customary manner, forming what is called the *diatonic* scale.

C		D		E		F		G		A		B		C′
1		9/8		5/4		4/3		3/2		5/3		15/8		2/1
	9/8		10/9		16/15		9/8		10/9		9/8		16/15	

The fundamental, major third, and fifth constitute what is known as a *major triad*, and the diatonic scale may be regarded as made up of three such triads. Thus starting the triad on C gives C, E and G, the triad on G gives G, B, D′ (where D′ is in the next octave), and finally the triad on F, gives F, A, C′. In each of these cases it will be found that the relative rates of vibration are 4 : 5 : 6.

The interval 9/8 is called a *large*, and 10/9 a *small whole tone*. The interval 16/15 is called a *half tone*, but this is larger than half of even a large tone, because, taken twice, it is greater than 9/8. In fact $(16/15)^2 = 1.1377$ while $9/8 = 1.125$. Interjecting tones midway between C and D, D and E, F and G, G and A, A and B, brings the scale much nearer to uniformity of small interval and makes it consist of twelve half tones, which are, however, not all equal. A scale derived in this way is called a *diatonic* scale.

377. Equal Tempered Scale. The above scale applied to instruments of fixed keys, as the piano, leads to confusion: when it is desired to go from one key as the starting point to another, the various irregular intervals will not occur in the proper place in the new series.

This difficulty finally led to the introduction of the *equal tempered scale* where all the half tone intervals are made equal to

the ratio of $\sqrt[12]{2}:1$ or $1.05946:1$. This ratio applied twelve times equals 2 to 1, and the octave is thus divided into twelve equal steps. Of course this temperament does not give true harmony and it is not customary to tune a piano by such a mathematical process. In practice the tuner establishes a number of notes throughout the keyboard and then proceeds to "distribute the wolf," that is, to tune the others in between so they will sound *as harmonious as possible* when played in various keys.

378. Perception of Sound Direction. The direction of a source of sound is deduced (usually unconsciously) from the difference of the effects at the two ears. When the source is near, its direction is shown by its sounding louder to the nearer ear. When a distant source gives a pure tone of high pitch (500 or more) and short wave-length and is to the right of the observer, the head throws a sound-shadow toward the left; and, since the left ear is in the shadow, it hears a fainter sound. There is no appreciable shadow when the waves are long compared with the dimensions of the head; but the phase at the right ear is ahead of that at the left, and Lord Rayleigh has shown that the observer unconsciously uses this difference to ascertain the position of the source. For pure tones between the above limits both processes play a part. In the case of complex sounds, the relative intensities of the components at the two ears will be different, since the components of shorter wave-length cast more distinct shadows. Hence the two ears will hear sounds of slightly different quality. If, in any case, the source is either directly ahead or behind, the head will have to be turned slightly before a decision can be made.

INTERFERENCE AND RESONANCE.

379. Interference of Sound Waves. Beats. Interference between two trains of sound waves of the same wave-length gives rise to places of no motion or nodes, and places of maximum motion or loops. The former are places of maximum variations of density, and an ear placed at such a point would hear a loud sound, while if placed at a loop, where there is no variation of density, it would hear no sound. Such effects are produced by the waves from the two prongs of a tuning fork. By turning the fork around near the ear, it is easy to distinguish places of no sound and places of loud sound. Similar effects are sometimes produced by a train of waves reflected from a wall interfering with the direct train, the source of the sound being an organ pipe.

Interference between sound waves of slightly different length is more common. It produces what are called beats, that is, a throbbing of the sound as it swells out and dies down alternately. This is the effect when two tuning forks, organ pipes or whistles of nearly the same pitch are sounded simultaneously. It is evident (§ 345) that the number of beats per second will be the difference of the frequencies of the interfering waves. When a single large tuning fork is moved rapidly toward a wall the ear hears beats. This is explained by Doppler's principle (§ 372). When the tuning fork is moving away from the ear the frequency of the direct sound heard is lowered. The frequency of the reflected sound heard is the same as that of the waves that fall on the wall toward which the tuning fork is moving, that is, it is raised. Hence beats ensue.

380. Resonance. The fact that air spaces respond with particular emphasis to tones of certain pitches is evidenced by the results of trying various tones in a cave, an empty room, or over the mouth of a cistern or similar empty vessel. A small impulse, if imparted at the right instant and oft repeated, may result in very considerable motion. The infinitesimal impulse imparted to a pendulum by the weight or spring of a clock would not move it appreciably, but the cumulative effect of many impulses delivered at the same frequency as that of the vibration of the pendulum causes it to swing through a considerable arc. A person walking on a spring board, a horse trotting or a procession marching upon a bridge may set it into dangerous vibrations. Figure 205 shows a section of a "spherical resonator." The small orifice *C* may be placed in the ear. Any impulse or sound-wave falling upon the opening *B* will enter the cavity *A*, be reflected from the rear walls and return to the opening *B*. If it arrives just in time to be reinforced by a second impulse at *B*, the air comes into exaggerated vibration and the tone is greatly strengthened or reinforced. The action is analogous to a closed organ pipe (§ 382), a node existing at the inner surface of the sphere *A*, and a vibration loop at the opening *B*. The pitch to which the air in any cavity responds depends upon the size and shape of the cavity *A* and the character

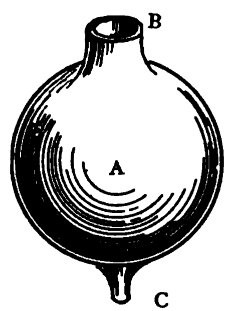


FIG. 205.

of the opening *B*. This effect is produced only when the natural period of the resonator is identical with that of the impressing tone. Such a resonator as is shown in Fig. 205 is closely "selective" and will not "respond" to tones even slightly removed from its own pitch. Such a resonator, when placed at the ear, will detect the presence or absence of its tone in the complex which is sounding. Helmholtz used such resonators to study the composition of complex sounds, especially to analyze vowel sounds. Resonance also plays an important role in determining the equality of a vowel as produced by the voice (§ 393).

VIBRATIONS OF BODIES.

381. Transverse Vibrations of Strings. The velocity with which a transverse wave travels along a string depends on the tension and mass of the string, not at all on the form of the wave (§ 353). If the tension be T dynes and the mass per unit length m in g. per cm.,

$$V = \sqrt{\frac{T}{m}}$$

Consider the motion of a stretched string set in vibration by a blow or by a bow applied at the middle *AB* of the string, Fig. 206.

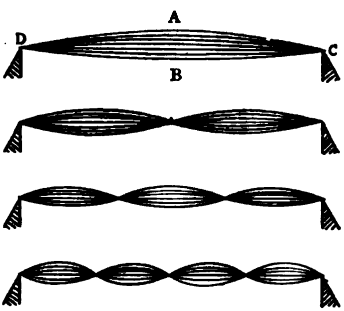


Fig. 206.

An impulse will be imparted to the string and this will travel along the string in both directions and be reflected from both ends successively. If the initial displacement is upward the reflected displacement will be downward and will meet at the middle of the string. Thus the string will have completed a half vibration and each disturbance will have traveled a distance, *ACB* or *ADB*, equal to the whole length of the string. These disturbances will continue to move in their respective directions and, after reflections at the ends of the string, they will again meet as upward displacements at the middle, when the string will

have completed one vibration and each disturbance will have traversed the length of the string twice, *i. e.*, *ACBDA* or *ADBCA*.

Now the length, λ, of a wave is the distance the impulse travels in one complete vibration. Hence $\lambda = 2l$. The velocity V of the impulse is (§ 381) $V = \sqrt{T/m}$ where T is the tension in dynes and m the mass of unit length of the string, and since $V = n\lambda$ (§ 349) where n is the number of vibrations per second it follows that

$$n = \frac{V}{\lambda} = \frac{1}{2l}\sqrt{\frac{T}{m}}.$$

Hence the rate of fundamental vibration of a stretched string is:—

(*a*) Inversely proportional to its length.

(*b*) Directly proportional to the square root of the tension.

(*c*) Inversely proportional to the square root of the mass of the string per unit length.

A vibration started in the way explained, by a simple impulse, will soon die away, as in the case of the piano string. To maintain them, successive impulses, synchronous with these free vibrations of the string, must be continuously applied. The resined bow applied to a violin string alternately impels the string in one direction and allows it to slip back in the opposite direction, so that the period of free vibration of the string controls automatically the frequency of the impulses. The pitch of the note is varied by shortening the effective length of the string with the finger.

Such a string may also perform partial vibrations, that is to say it may in vibrating divide up into two halves, three thirds and so on as indicated in the figure. We may suppose such vibrations to be produced by impulses at double, treble, etc., the rate of the former, applied at proper points. A point of percussion cannot become a node. Each vibrating section of the string may be regarded as a separate short string, and to find the frequency we may apply the above formula, substituting for l the length of the part; that is, when the string vibrates as two halves we substitute $l/2$ for l, when as three thirds, we substitute $l/3$ for l and so on. Hence, if the frequency of the vibration of the string as a whole is n, the frequencies of the partial vibrations are $2n$, $3n$, etc. That is to say the frequencies are as $1:2:3:4$, etc.

These different modes of vibration have been described as taking place separately, nevertheless they may all exist simultaneously.

In fact, when a piano string is struck at any point, an irregular disturbance is produced, which may be considered as a complex harmonic motion, consisting of numerous simple harmonic elements of frequencies proportional to 1:2:3, etc. (Fourier's theorem, § 352). Each element is propagated along the string in the manner described above and establishes a set of stationary waves with the frequencies stated, except that evidently no set will arise which requires the point of percussion to be a node. This last statement enables us to understand why the quality of the note produced by a string depends on the point at which it is struck, plucked or stroked, the most pleasing effect being produced when the point is about one-eighth of the length of the string from one end.

When any source of a musical sound, such as a string, pipe, or fork, is capable of vibrating continuously in more than one steady state, the tone produced while it is vibrating in its slowest and simplest mode is called its **fundamental**, and the tones due to the other modes are called **overtones** or **partials**. If these latter have frequencies two, three, etc., times that of the fundamental, they are called **harmonics**.

382. Vibrating Columns of Air. The simplest mode of vibration of a column of air in a pipe **closed** at one end is shown (on a greatly exaggerated scale) in Fig. 207, in which the spacing of the

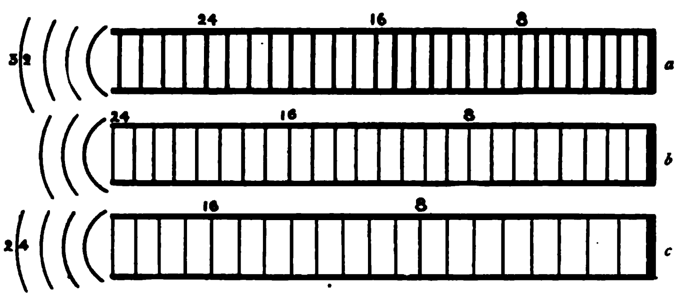


FIG. 207.

numbered lines indicates degrees of condensation or rarefaction. The air rushes into the open end and compresses itself against the closed end, as shown by *a*. This condensation relieves itself through the open end and its inertia carries it beyond the neutral condition, *b*, to the other extreme of rarefaction *c*, and the opera-

tion repeats itself periodically. This steady state of vibration of air in a closed pipe is diagrammatically represented by *a,* Fig. 208. The train of waves entering the open end of a pipe of length L and reflected back from the closed end combine to form a stationary wave system (§ 355), of which, in the simplest case, there is a quarter wave length, $\lambda/4$, in the pipe, that is the wave length is $4L$. The first overtone or next steady state is represented by Fig. 208*b*, where the waves are shorter and there is three quarters of a wave in the pipe, or the wave length is $4L/3$. The third steady state is shown at Fig. 208*c*.

Similarly if the pipe is **open** at both ends and a compression

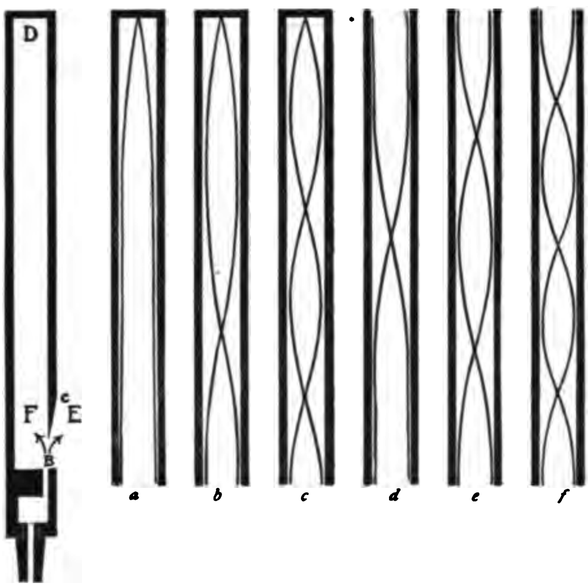


FIG. 208.

enters each end simultaneously they will meet in the center and conditions will be as if they were each reflected back. In reality they pass through each other and, by repeated reflections at the open ends of the pipe, without loss of phase, form a system of stationary wave with a node at the center, Fig. 208*d*. One half a wave is in the pipe, hence the wave length is $2L$. The first over-

tone is shown at Fig. 208*e* with a complete wave in the pipe so that the wave length is *L*. Fig. 208*f* is the third overtone.

383. Closed Organ Pipe. In the closed organ pipe, the air issues in a thin flat jet from slit *B* (Fig. 208), and impinges on the edge *C* of the wood or metal of the pipe. The jet starts by splitting on the edge *C* into two parts *E* and *F*. Each part produces a slight condensation of air or compression which travels off with the velocity of sound. The outside impulse *E* is dissipated, but the inside impulse *F* travels up the pipe. Thus we have the condition required for stationary vibrations in a pipe closed at one end as described in § 382. These vibrations would of course soon die away unless energy were continually supplied to make up for that lost in friction. This energy comes from the energy of the stream of air directed against *C*. The jet vibrates backward and forward in unison with the vibrating air-column and thus continues to deliver energy to the vibrating air.

We may also regard the pipe as acting as a resonater (§ 380). The breaking of the jet of air on *C* produces a complex sound from which the tube selects that component with which it is in unison and stationary vibrations are thereby set up.

As stated in § 382, the wave length of the fundamental tone of the pipe is four times the length of the pipe, though minor influences prevent this from being absolutely true. The pipe may also sound in a steady state when the wave length is (as in Fig. 208*b* and *c*) 4/3 the length of the pipe, 4/5 the length and so on. It is thus seen that the wave lengths shorten in the ratios 1 : 1/3 : 1/5. etc., and hence the frequencies of the harmonics or overtones produced by a closed pipe rise in the ratios of the odd number 1 : 3 : 5, etc., or if the pipe is of such a length that its fundamental is C the successive tones are C, G′, E″, etc. (see § 377).

384. Open Organ Pipe. The essential difference between the open and the closed organ pipe is the fact that an impulse traveling in the confined space of the pipe arrives at an open end and finding freer conditions is reflected as if from a less dense medium without loss of phase. That is to say the individual displacements in the incident and in the reflected wave are in the *same direction*. Under these conditions of reflection the condensation is *practically* reflected as a refraction and vice versa (§ 354). From *d, e, f* of Fig. 208 it is evident that the wave length of the fundamental

of an open pipe of length L is $2L$, that of the first overtone is $2L/2$, that of the second overtone $2L/3$, and so on. Hence the frequencies are as $1:2:3$, etc. Comparing open and closed pipes of the same length L, it is seen that the wave lengths of the fundamentals are $2L$ and $4L$, or in the ratio of $1:2$, that is to say, a pipe lowers its pitch an octave on being closed. Again minor influences render this statement not exactly accurate.

385. Longitudinal Vibration of Rods. When a rod is clamped at one end and free at the other, it is capable of longitudinal vibrations like the air in a closed organ pipe. If the rod is struck on the end or stroked lengthwise with a rosin-covered cloth, it may be made to emit a musical sound. The pitch is determined by the length of the rod and the velocity of propagation of longitudinal waves, in the material of the rod. As in the closed organ pipe the length of the rod equals one fourth the wave-length of the fundamental vibration. A rod clamped at the middle is similar to an open organ pipe and its length equals one half the wave-length of the fundamental vibration in the material. By determining the pitch or number of vibrations per second of such a rod, the velocity of propagation of the waves in that material may be calculated, since $v = n\lambda$, where v is the velocity, n the rate, and λ the wave length or twice the length of the rod. In this case the wave is one of linear extension and contraction, and the modulus of elasticity of such strains is the stretch-modulus (§ 171). Hence an expression for the velocity may be deduced from that of § 365 by substituting M for E, *i. e.*, $v = \sqrt{M/\rho}$ where M is Young's modulus, and ρ is the density of the rod. (In the case of a solid the isothermal and adiabatic moduli are practically equal.)

386. Kundt's Method of Finding the Velocity of Sound. This is a method by which the velocity of sound in a gas can be found when the velocity of longitudinal waves in a rod is known or *vice versa.*

A glass tube 4 or 5 feet long AA is clamped at B on a table T. The rod DD is supported at E and carries on its inner end a disk which almost fills the cross section of the tube. C is a gas-tight piston which may be moved back and forth in A. A small quantity of cork filings are placed along the inside of the tube. By stroking the outer end of the rod D it will be put into longitudinal

vibrations, and the disk F will set the air in A into a set of stationary waves. The to and fro motion of the air will cause the cork particles to arrange themselves in ridges at vibration loops

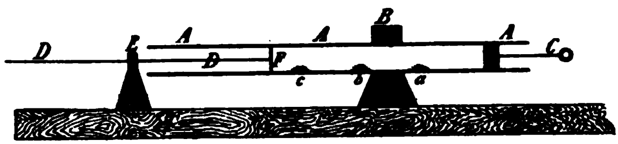


FIG. 209.

as at c, b, a. C is adjusted so that the action is greatest and the ridges well defined.

The length of the wave in the rod is $2L$ where L is the length of the rod, and the length of the wave in air is $2l$ where l is the mean of the distances ab, bc, etc. The frequency of vibration, n, is the same for both. Hence denoting the velocity of longitudinal waves in the rod by V and the velocity of sound in air by v

$$n = \frac{V}{2L} = \frac{v}{2l} \quad \text{and} \quad \frac{V}{v} = \frac{L}{l}$$

This gives us V since v can be found by other means (§ 363) and then by putting different gas in the tube AA or varying the temperature and pressure conditions the velocity in various gases under various conditions may be investigated.

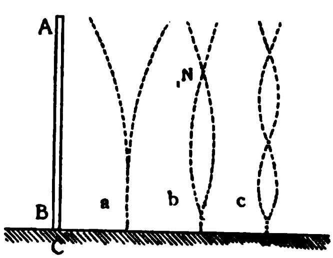


FIG. 210.

387. Transverse Vibrations of Rods. If a thin rod AB is firmly fixed at one end as shown in Fig. 210 it can vibrate in a variety of modes. Thus a represents its fundamental mode of vibration and b and c its first and second overtones. The frequencies of the modes are approximately as $1 : 6.25 : 17.5$.

388. Tuning Forks. If the rod AB, Fig. 211b, be bent as indicated successively in b, c, d, the nodes N_1 and N_2 will gradually approach each other until in d are represented the conditions in a tuning fork, where the nodes are at the bottom of the prongs near

the shank. By considering the modes of the over tones of *AB* shown at *ef* and *gh*, it will be readily seen that the over-tones of

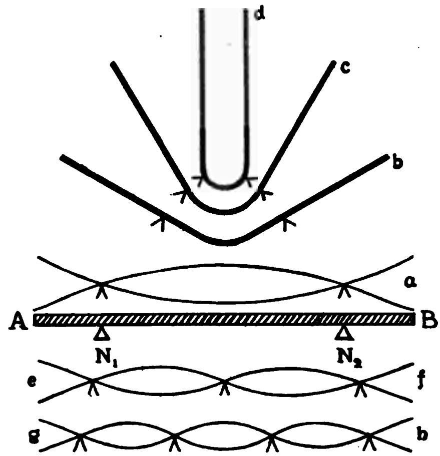


FIG. 211.

a fork are very high and not in harmony with the fundamental; but they are not readily started and die away much more rapidly

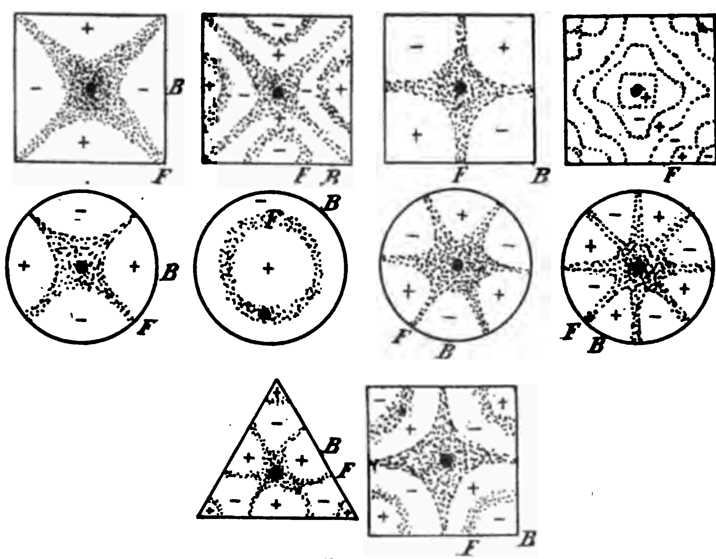


FIG. 212.

than the fundamental, and, for this reason, tuning forks are particularly adapted to furnish pure tones.

389. Vibrations of Plates and Bells. Many of the experimental facts of vibrating rods, etc., are due to Chladni (who died 1827). He particularly investigated the modes of vibration of thin plates of various shapes. When a square plate of glass, brass, aluminum, or the like, is clamped at its center, it is capable of steady vibration in various modes. These can be determined by the point, *F*, at which the finger is held to establish a nodal line, and the point *B*, at which the violin bow is drawn across the edge of the plate to establish a vibration loop. If the plate is horizontal, sand strewn upon it will collect along the nodal lines, being thrown off the loops. Fig. 212 shows a variety of modes for square, circular and triangular plates, the black spot being the point of support. The sign + indicates that at a given instant these parts of the plate are above the neutral plane while those marked — are below it.

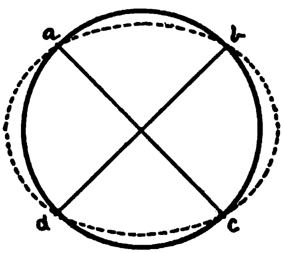


FIG. 213.

390. Bells. The vibration of a bell may, to a certain extent, be considered as analogous to the vibration of a circular plate. When sounding its fundamental tone a bell vibrates with four nodes at *a b c d* (Fig. 213).

391. Reeds. In certain organ pipes the original impulse is given, not by the impigning of a jet of air, but by stopping and starting it by the vibrations of a metal strip. The thin strip of metal

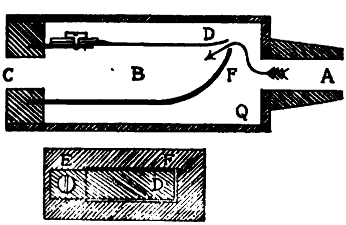


FIG. 214.

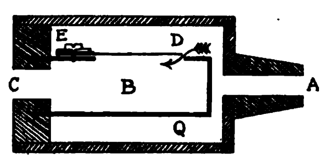


FIG. 215.

D (Fig. 215), screwed to the frame *F*, just swings clear in the rectangular opening of the plate *F*. The air, entering through

A, impinges on the reed *D*, and sets it into vibration. *D* in its motion alternately opens and closes the opening between itself and *F* and thereby allows a series of puffs of air to pass to *B* and *C*. The length and elasticity of the strip determines the rate of the reed and the pitch of the tone produced. Such reeds, called *free reeds*, are used in the parlor organ and melodion, and the familiar mouth organ.

If the reed is long enough and wide enough, it will cover the opening in the frame *F* (Fig. 214), and in its vibrations it will strike the frame *F*. In this case it is called a *striking reed*. Here again the pitch is determined by the dimensions of the reed. Such reeds are most familiar in the common tin-horn.

392. Reed Quality. The puffs of air issuing from reed pipes are thus controlled by a mechanical shutter and are not simple in their nature; hence their quality is very complex and the tone appears "nasal." If a pipe is attached to the opening *C* in either form, it will considerably affect the quality of the sound, since by its natural resonance it will favor certain tones and discourage others. Those elements in the complex tone of the reed which coincide with the natural rates of the pipe will be made to predominate. In fact the pipe may to some extent impose its rate upon the reed as is the case in the clarionet and the oboe. In practice, where reed pipes are used in organs, the reed and pipe are tuned to the same fundamental pitch.


FIG. 216.

The flute is an open lip-pipe where the form of vibration and the frequency are determined by certain openings along the side of the tube. The clarionet, cornet, etc., may be classed as reed-pipes. In the trombone the pitch is controlled by varying the length of the pipe.

393. The Voice. The *pitch* of the tone produced by the voice is determined by the length, tension and weight of the vocal cords (shown wide open in Fig. 216) which are set in vibration by the exhaled air, much as is the reed in the reed pipe. The *quality* of the tone is controlled by the size, shape and opening of the cavities of the mouth and nose. These can be altered in dimensions so as to resonate to different components of the complex vibrations of

the vocal cords (§ 369). One vowel differs from another only in the *quality* of the sound as determined by the resonance cavities of the mouth and nose. Consonants are merely the particular way of starting or stopping a vowel sound.

394. The Ear. The complete apparatus by which man hears consists of four distinct parts. First, the ***external ear,*** which serves to collect the sound waves and conduct them down the small tube to the ***drum*** of the ear which is set into vibration by the vibrations of the air. The motions of the drum are transmitted to the liquid in the ***vestibulum*** by the intervention of three little bones. The vestibulum is filled with a watery liquid and lined with nerve tendrils, which stand out toward the center like hairs in fur. This portion of the ear is undoubtedly intended to recognize sound as such and independent of pitch. Extending beyond the vestibulum and connecting with it is the *cochlea* or snail shell. This is a long conical tube wound up in a spiral like a snail shell, as its name implies. Across the interior of this tube are stretched some 3000 nerve tendrils called Corti's fibers from their discoverer. Each little Corti nerve fiber is attuned to some particular pitch and immediately conveys to the brain a like pitch caused by the complex motion of the liquid. From the relative intensity of the reports of the nearest fibers the brain locates the pitch of a tone falling in pitch between two fibers.

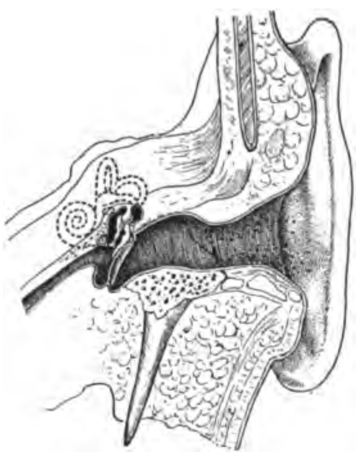


FIG. 217.

It is this harp of three thousand strings which enables one to judge of pitch and quality.

REFERENCES.

RAYLEIGH'S *Sound,* the standard work of reference. It presumes thorough acquaintance with mathematical analysis.
POYNTING & THOMPSON'S *Sound.*
BASSET'S *Elementary Treatise on Hydrodynamics and Sound.*
HELMHOLTZ'S *Sensations of Tone.*
BARTON'S *Text-book on Sound.*

PROBLEMS.

Velocity of sound in gases. 1. What would be the velocity of sound in a gas consisting of 5 parts hydrogen, 3 parts oxygen, and 2 parts nitrogen, at 76 cm. and 0° ? Ans. 45,017 cm./sec.

2. A rifle is fired at a target distant 1220 meters. Assume the velocity of the bullet to be 800 meters per second, and the air dry and at 20° C. (*a*) At what distance on the line joining gun and target would an observer hear the report of the gun and the impact of the bullet on the target at the same instant? (*b*) On a line through the target perpendicular to the line joining gun and target, at what distance from the target would an observer hear the report and impact simultaneously ? (*a*) 357 m. (*b*) 1227 m.

3. After a flash of lightning three seconds elapse before the first sound is heard ; how far distant is the discharge ? 990 m.

4. A lightning discharge takes place somewhere between a certain horizontal cloud and the earth. The direct sound is heard in two seconds and the echo from the cloud is five seconds. How high is the cloud and how high is the discharge ? 1155 m. 660 m.

5. A stone is dropped from a balloon and is three times as long falling to the ground as the sound of the impact is in returning to the balloon; how high is the balloon ? 2469 m.

6. A cliff distant 500 feet from a source of sound gives back an echo in 0.918 second, as measured by a chronograph. What value for the velocity of sound will this give in cm. per second ? 33,202 cm./sec.

7. The frequency of a certain musical note is 256 per second, and its wave-length in a certain gas is 125 cm. What is the velocity of sound in that gas ? 32,000 cm./sec.

8. The frequencies of the highest and of the lowest note that can be heard are 40,000 per second and 20 per second respectively. Taking the velocity of sound as 34,000 cm. per second, what will be the wave-length in air of the waves corresponding to these notes ? .85 cm. 1700 cm.

Doppler's Principle. 9. Show how the pitch of a note is changed by a motion of approach or recession on the part of either the observer or the sounding body.

10. A whistle on a train has a frequency of 500 vibrations per second. If the train is moving at the rate of 30 miles per hour, what is the change in pitch as the train passes an observer, assuming the velocity of sound to be 1120 ft./sec.? 39.2 vib./sec.

Musical scales. 11. Assume pitch *a* to have a rate of vibration of 437, calculate the rates of the other notes of the same octave in the diatonic scale, and in the equal tempered scale.

12. A tuning fork *A* corresponds to rate of rotation of a syren of 220. in 10 seconds; the disk has 16 holes. What rate of rotation would produce the fifth above this pitch ? 33 rev./sec.

Vibration of bodies. 13. How may a segment length in a stationary wave-form be proved equal to half a wave-length ? In the transverse vibration of a cord, show the frequency of the fundamental and the first three upper partials, in terms of the length of cord and velocity of waves in the cord.

14. An aluminum wire, 1 mm. in diameter is stretched between two supports, 1 meter apart, with a tension of 4 kg. What is the pitch of the fundamental note given when it vibrates transversely ? 68.4 vib./sec.

15. Two strings of equal length, 80 cm., are of brass, and are under the same tension. One string is 0.12 mm., the other 0.13 mm. in diameter. The stretching weight is 10 kilograms. How many beats per second are produced by the two strings sounding together? 48.2 beats per sec.

16. What will be the pitch of the fundamental note given out by an organ pipe, closed at one end, and 30 cm. long? Velocity of sound is 34,000 cm./sec. What will be the pitch of the first overtone? 283.3 vib./sec. 850 vib./sec.

17. The distance between the dust piles in a Kundt's tube was 10 cm. The rod used was 1 meter long and clamped at its middle point. If the tube was filled with air at 20° C., what was the velocity of sound in the rod? 3427 m./sec.

18. A metal rod 125 cm. long when held at the center and struck on the end gives a note of 1,000 vibrations per second. If the density of the material is 7, what is the value of Young's modulus? $.437 \times 10^{12}$.

19. Indicate the nodal lines upon a square plate when vibrating to the third lowest note of which it is capable. Mark by + the parts above normal, and by — those below.

LIGHT.

By E. Percival Lewis, Ph.D.

Professor of Physics in the University of California.

GENERAL PROPERTIES.

395. Radiation. As in the cases of Heat and Sound, the word Light has acquired two distinct meanings. The primary and more familiar one is that which is associated with the sensation of vision. Nearly all that relates to this aspect of the subject lies within the province of the psychologist. The physicist, however, generally uses the term in an objective sense, with reference to the external agencies which may excite the sensation of light if allowed to act on the eye. The visible radiation which affects a normal eye will also affect a photographic plate, a thermometer, or other sensitive detector of heat. It will be found, after analyzing the radiation from the sun, electric light, or other sources with a prism, that beyond the violet and the red lie non-luminous radiations which will affect a photographic plate or a thermometer, and it will also be shown that the oscillations of an electric spark between metallic terminals are accompanied by the radiation of electric waves through space. There is, as we shall see, no fundamental qualitative difference between these various radiations, and it is due merely to an accidental property of the eye that some of them excite the sensation of light while others do not. This is analogous to the selective resonance of a piano wire, which will respond to certain notes and not to others. Just as some ears can detect sounds of such high pitch as to be inaudible to others, some eyes can detect ether radiations lying somewhat beyond the limits of perception of the ordinary eye. In the following pages the whole range of these radiations so far as they are known will be considered. As a matter of convenience, the term **Light,** which strictly speaking would apply only to the radiations exciting the sensation of light, will be used in a figurative

sense to include the entire range of radiations which are alike in their general properties, and which were once very artificially classified as luminous, actinic, and heat radiations.

The term **Radiation**, which means the emission of energy through space from any source, may be applied to both Sound and Light phenomena. To distinguish them, we may use the term *Sound Radiation*, which applies to a form of radiation through the medium of matter, and *Ether Radiation*, which, as shown by the passage of light through celestial space, is propagated independently of matter, although in many cases modified by its presence.

Other very interesting and important classes of radiation have recently been brought to our knowledge, such as the Becquerel rays, which appear to be due in part to the projection of small particles from such substances as uranium, thorium, and radium compounds, and Röntgen rays, which seem to be a form of ether disturbance; but the mode of production and of propagation of these radiations is very different from that of light (see Radioactivity).

396. Sources of Light. The best known are the sun, the physical nature and condition of which are as yet not fully understood, solid bodies at a high temperature, such as the calcium light, electric arc and incandescent lights, and luminous flames. If a piece of cold porcelain is held over the flame of a candle, lamp, or gas jet, it will become covered with finely-divided carbon, while no such deposit is observed in the case of a non-luminous bunsen or alcohol flame. This suggests that the luminosity of these flames is due to the presence of incandescent carbon particles. This idea is strengthened by the fact that when the base of a bunsen burner is closed the flame becomes luminous and smoky; when open, enough oxygen is admitted to combine with all the carbon set free by the dissociation of the coal gas, and the flame is then non-luminous. The carbon oxides formed are permanent gases, and there is no evidence that such gases can be made luminous by high temperature alone. Any gas may be made luminous, however, by the passage of an electric discharge through it, but this luminosity does not seem to be accompanied by very high temperature. There are, in fact, many cases in which light is emitted at a very low average temperature of the source. As examples may be mentioned the various types of phosphorescence, some of which are most active at temperatures as low as that of liquid air, the

aurora, due to electrical discharges through the highly rarefied and very cold upper atmosphere, and the light emitted by fireflies and glow-worms.

397. Rectilinear Propagation. One of the earliest observations concerning light was that it travels in straight lines, in a homogeneous medium. These lines of propagation or "rays" may be made to alter their direction only by one of two methods—by reflection, when they fall on the boundary between two media, or by refraction, when they pass obliquely from one medium to another, or through a medium of varying density.

398. Shadows and Eclipses. Rays pass in straight lines by the edges of an obstacle, so that the space behind it is screened from the light. If the latter comes from a very small or "point" source the shadow would be sharply defined if the propagation were strictly rectilinear; as a matter of fact, close observation shows in all cases that the light fades gradually into the shadow. This very significant fact proves that light travels only approximately in straight lines; there is always more or less lateral spreading. Strictly speaking, there is, then, no such thing as a ray of light, if we mean by this term propagation along a geometrical line. The explanation of this spreading will be given later (§ 420).

A more obvious cause of the lack of sharpness in shadows is to be found in the fact that most sources of light are not even approximately points, but are of finite area. This gives rise to the distribution of light and shadow shown in Fig. 218 and Fig. 219. The first represents the shadow cast by an object larger than the source; the second, that due to an object smaller than the source; for example the shadow of the earth due to the sun. In each case there is a region of complete shadow behind the obstacle, called the umbra, into which no light from any part of the source can enter. Around this there is a region called the penumbra, which receives light from a part of the source, the effective portion of the latter increasing in going outward from the umbra. When the moon lies entirely within the shadow cone of the earth it is said to be completely eclipsed; when it passes through the

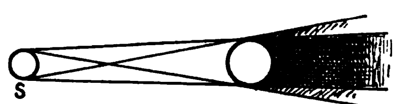


FIG. 218.

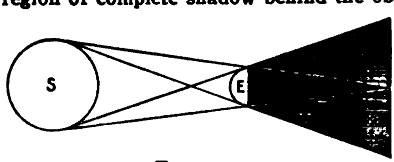


FIG. 219.

penumbra or partly through the umbra and partly through the penumbra it is partially eclipsed.

399. Parallax. This well-known phenomenon depends upon the rectilinear propagation of light. By parallax is meant the apparent displacement of an object due to the real displacement of the observer. For example, if the observer moves from O_1 to O_2 (Fig. 220) A will appear to be displaced an angular distance $\alpha + \beta$ to the left with reference to B. That object which seems to be displaced in a direction opposite to the motion of the observer is evidently the nearer. When traveling on a railroad train objects near at hand appear to be moving backward, those at a distance in the same direction as the observer. If two objects are coincident in position or equally distant their relative parallax vanishes. This gives a useful method of finding the apparent position of the image formed by a lens or mirror, or of focusing the cross thread of a telescope. When the latter and the image of a distant object are both distinctly seen and have no relative parallax they are coincident in position and both in focus.

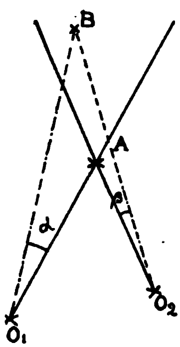


FIG. 220.

In astronomy horizontal parallax is defined as the angle subtended by the semi-diameter of the earth from any body of the solar system. Annual parallax is the angle subtended by the semi-diameter of the earth's orbit from the more distant fixed stars. The distance between the sun and the earth may be determined by observing the transit of an inferior planet, Venus for example, across the sun's disc. Observers at A and B (Fig. 221) note the instants at which Venus appears to enter the sun's disc as viewed from their respective stations. From the interval between these two contacts and the known angular velocity of Venus around the sun the angle α may be determined, and from that and the base line AB the horizontal parallax of the sun may be calculated. Of course correction must be made for the motion of the earth between the instants of contact.

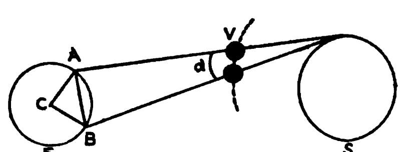


FIG. 221.

400. Pinhole Image. Another effect of the approximately recti-

linear propagation of light is the formation of an inverted image of the source by light passing through a small orifice such as a pinhole. If any source, for example, a candle, is placed opposite such a hole in a screen S_1 (Fig. 222) light from the point P will pass through the opening in a narrow cone or *pencil* and illuminate a small patch at P_2 on a screen S_2. Light from Q will form a small patch at Q_2, and light from any other point of the flame will fall on a corresponding point of the screen S_2. The group of patches will in form, color, and relative brightness reproduce the candle flame, but evidently inverted in position. The pinhole forms an

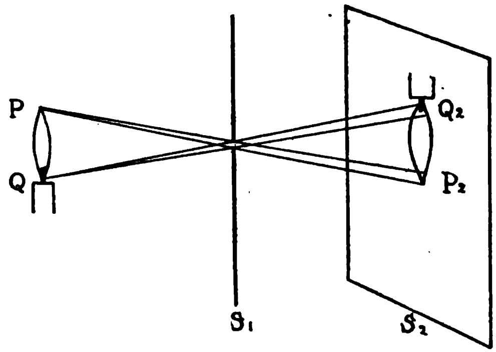


FIG. 222.

image like that due to a condensing lens, but the total light in the pinhole image will be less than that formed by the lens in the proportion of the area of the pinhole to that of the lens. As the image is due to a group of overlapping patches, it will not be so sharp in outline as that made by the lens. The blurring will increase with the size of the opening or when the source is brought near the screen, thus increasing the angle of the transmitted cone. The object and its image subtend equal angles at the pinhole, so that their linear magnitudes are in the same ratio as their respective distances u and v from the screen S_1. This is also true of images formed by any optical device, such as a mirror or lens.

Pinhole images were first observed by the great Italian artist and scientist, Leonardo da Vinci, about 1500. The camera obscura, in which the image is improved in sharpness and brightness by the use of a lens, was

invented by Porta, an Italian physician, about 1589. The ordinary photographic camera represents the final stage of improvement of this device. Landscape photographs of great softness and beauty may be made by the use of the pinhole camera.

401. Reflection, Regular and Diffuse. When light falls on a smooth polished surface it is reflected in a definite direction. This is called regular reflection. The plane including the direction of the incident light and the normal to the surface at the point of incidence is called the *plane of incidence*. The angle between the incident pencil and the normal to the surface is called the *angle of incidence;* that between the reflected pencil and the normal is called the *angle of reflection*. Experiment shows that (1) *the angle of reflection is equal to the angle of incidence;* (2) *the reflected pencil lies in the plane of incidence*. It is evident from the first law that if a mirror is rotated through a given angle the reflected pencil will be rotated through twice that angle.

When light falls on a rough unpolished surface it is reflected in all directions. This is called diffuse or irregular reflection. There is no essential difference between regular and diffuse reflection except that in the latter we may imagine reflection to take place from an infinite number of infinitesimal plane surfaces orientated in all directions.

402. Visibility of Objects. On a clear night, when there is no moonlight, the stars and planets appear against a background of black sky. The space around the earth's shadow cone is filled with sunlight, but we do not see it unless it is reflected from some planet or the moon. If a beam of light is passed through a vessel of distilled water its path is invisible. If a beam of sunlight enters a dark room it cannot be seen unless dust particles are floating in the air. A drop of milk in the water or a little dust stirred up in the room will cause the path of the light to flash out brilliantly. Such experiments show, as might be expected, that light does not excite the sensation of luminosity unless it enters the eye directly from the source or by reflection. Ordinary objects are visible because they reflect light diffusely into the eye, and they may be regarded as secondary sources of radiation. A perfect reflector would itself be invisible, all the light reflected from it appearing to come from the image of the source, not from the reflector.

403. Transmission and Absorption. Light travels through some media, for example most gases, glass, and water, with scarcely any appreciable diminution of intensity. Other media may transmit little or none, or certain colors only; such media are said to show general or selective absorption. In cases where absorption occurs there appears to be a loss of radiant energy, but it may be shown that it is changed to other forms, usually heat (§ 309 *et seq.*).

404. Transparency, Translucency, Opacity. Any substance which transmits a large fraction of the incident light without scattering it is said to be transparent. As indicated by this term, objects may be seen clearly through such substances. Objects which absorb all the unreflected incident light are said to be opaque, and act as perfect screens. Evidently any perfect reflector must also be perfectly opaque, but in this case opacity is not due to absorption. Substances differ widely in these properties, varying from almost perfect transparency to almost perfect opacity. The most transparent media known show some absorption, which increases with the length of path; hence any substance will become opaque if a sufficient thickness is taken.

No light penetrates to great depths in the ocean, although a layer of water of considerable thickness is transparent. On the other hand, light will penetrate to a slight depth in any medium, so that thin layers of metal or of carbon are found to be transparent. Some substances are selectively transparent; red glass will freely transmit red light, but not the other colors, and a thin sheet of hard rubber, which appears to be opaque, will transmit radiations lying a little outside the red of the spectrum (§ 311).

Some substances transmit light, but scatter it so that objects cannot be clearly seen through them. These substances are called translucent. The effect is caused by diffused reflection within the medium, due to discontinuity or non-homogeneity of structure, as in the case of powdered glass, paper, or water containing finely-divided particles. Some substances, such as paraffin, are homogeneous and transparent when in the fluid state, and translucent when in the solid state. The latter effect is apparently due to granulation or crystallization.

405. Refraction. When light passes obliquely from one transparent medium to another a part is usually reflected, while that which enters the second medium changes its direction abruptly at the boundary. Generally (but not always) in passing from a lighter medium to a denser the light is deflected toward the normal to the boundary. This is called refraction. Since objects appear

to be in the direction from which the light comes, refraction, by changing the course of the light, causes an apparent displacement of the source. An example is found in the classic experiment of Kleomedes, who showed that a coin placed in the bottom of a vessel so that it is barely concealed by the sides of the latter, is apparently lifted into view when the vessel is filled with water (Fig. 223). The object at A then seems to be at the point A', and the bottom PQ of the vessel appears to be raised to $P'Q'$. Similarly, a meter rod dipped obliquely into water appears to be bent, and the divisions seem to be shortened. The latter effect is also observed when the rod is normal to the surface. This change in the apparent distance of objects seen normally through a refractive medium is to be considered as an example of refraction, although there is no deviation of the light. It will be shown in § 442 that these effects are the result of differences of velocity of light in the media concerned.

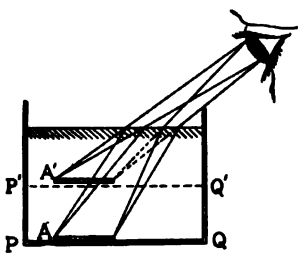


FIG. 223.

406. Intensity of Light. The brightness of light as estimated by the eye is not capable of precise physical determination. It depends to a large extent upon the color of the light and the sensitiveness of the eye. The only consistent way in which intensity of radiation may be determined or expressed is in terms of energy.

If radiation travels through a homogeneous medium in straight lines, and if the medium is perfectly transparent and does not itself emit radiation, the same total amount of energy must flow per second through any spherical surface concentric with the source. It follows that the *intensity* or quantity of energy passing through unit area per second, must vary inversely as the square of the distance from the source (§ 361).

The above conclusion is based upon the assumption that the radiation diverges uniformly in straight lines in all directions. It is not true if the medium is of varying refractivity, on account of partial reflection and of changing divergence of a cone of light in passing from one medium to another. In case a beam is made parallel by a lens or mirror there is no

change of intensity with distance except that due to absorption or to imperfect parallelism.

407. Photometry. The eye can form no exact estimate of degrees of intensity, but it can determine with great accuracy whether two adjacent surfaces are equally illuminated by lights of the same color. Upon this principle are based the different methods of **Photometry** or comparison of intensities. Two of the simplest and oldest types of photometer are the Rumford shadow and the Bunsen grease spot photometers. In the use of both it is assumed that the light from the two sources compared contains the different colors in the same proportions, making comparison possible.

In the photometer devised by Benjamin Rumford shadows of a rod R are cast on a white screen by the sources S_1 and S_2 (Fig. 224), one of which is a standard comparison source. By adjusting the positions and distances of S_1 and S_2 the shadows may be made to touch and to be of equal intensity. When this is the case, it is evident that the intensity of light from each source is the same at the screen, since each shadow is illuminated solely by the source which casts the other shadow. If this intensity is I, and if the intensities of the sources at unit distance are respectively I_1 and I_2,

$$I = \frac{I_1}{r_1^2} = \frac{I_2}{r_2^2}$$

or the respective intensities of the two sources are directly as the squares of their distances from the screen when the latter is equally illuminated by both. This relation holds likewise in the use of the other forms of photometer described below.

The Bunsen photometer consists essentially of a grease spot on a screen of white paper. Such a spot is more translucent than the clean paper, and for this reason appears darker by reflected light (since there is less light reflected from the spot). If such a screen is placed between sources which equally illuminate it with light of the same quality (same proportions of different colors) the grease spot will disappear. The loss in light reflected from the spot on one side will then be compensated by the increased amount transmitted from the other side.

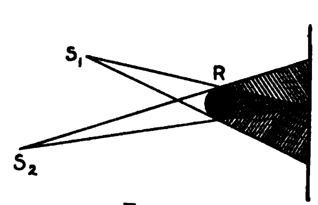


FIG. 224.

The Joly diffusion photometer consists of two rectangular blocks of paraffin separated by a piece of tin foil. Paraffin is a translucent substance which appears to scatter light throughout its entire mass. If this photometer is placed between two sources of light with the tin foil at right angles to the line joining them each block will be illuminated by one source alone. If the intensity of illumination is the same on both sides the boundary line between the two blocks will disappear; if it is not the same, the boundary is clearly seen, the block receiving the smaller amount of light appearing darker than the other throughout its entire mass.

408. Lambert's Law. A flat flame or an incandescent sheet of metal appears to be equally bright whether viewed normally or obliquely to its surface. The intensity or the energy falling per second on unit area of the surface BC (Fig. 225) is equal to the total energy emitted per second from AB at the angle α with the normal to the surface, divided by BC, or, if E_α is the emissivity of AB per unit area in that direction, $E = E_\alpha(AB/BC)$. The normal emissivity is E_n and observation shows that $E = E_n$: therefore

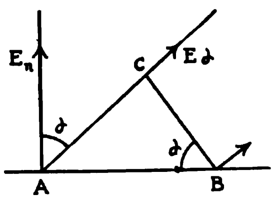


FIG. 225.

$$E_n = E = E_\alpha(AB/BC) \text{ or } E_\alpha = E_n \cos \alpha$$

This is known as Lambert's Law. In accordance with this principle, an incandescent sphere when viewed from a distance appears to be a uniformly illuminated disc.

The law does not apply to a surface bounded by an absorbing atmosphere, which will of course exercise greater total absorption in an oblique than in a normal direction. The sun, for example, which is surrounded by an absorbing atmosphere of gases, appears (as clearly shown in photographs) to be darker at the edges than at the center.

In the same way it may be shown that if I_n is the intensity of light falling normally on a screen, the illumination when the light is incident at angle i is

$$I_i = I_n \cos i$$

VELOCITY OF LIGHT.

409. Velocity of Light. The sensation of light is produced by a disturbance originating in distant bodies, and it may naturally be assumed that this disturbance travels with a finite velocity. Galileo, about 1600, appears to have been first to attempt to measure this velocity.

His method was substantially the same as that ordinarily used to determine the velocity of sound. Two observers stationed at some distance from each other endeavored to note the instants at which flashes of light from one station were observed at the other. The failure of such attempts made it clear that the velocity of light is so great that the time required to pass over ordinary distances is too small to be measured except by methods much more refined than those at that time available. It was natural, therefore, that the first results should have been obtained by astronomical methods, in which the distances employed are those between heavenly bodies.

In 1675 Olaus **Römer**, a Danish astronomer, observed that the eclipses of Jupiter's satellites by that planet recur at regularly increasing or decreasing intervals, according to the earth's position with respect to Jupiter. If the first observations are made when Jupiter and the earth are in conjunction, or on the same side of the sun and in line with it, the interval between the first and the second eclipse of one satellite is about 1 day 18.5 hours, but as the earth proceeds in its orbit the interval between eclipses slowly increases, so that at opposition, when the earth is on the opposite side of the sun from Jupiter, the eclipse occurs about 16 minutes later than the time calculated from the first observed interval. As the earth continues on its course the intervals diminish, the original value recurring at conjunction. During the period between conjunctions 225 eclipses of this satellite occur. The interval between the first and the 113th is greater by 16 minutes 41.6 seconds (according to modern observations) than that between the 113th and the 225th.

Römer explained this as being due to the finite velocity of light, the last instalment of light which comes from the satellite before the eclipse at opposition having to travel a distance equal to the diameter of the earth's orbit, in addition to the distance between the earth and Jupiter at conjunction, and in consequence reaching the earth later than if the latter had remained at the original distance from Jupiter. If n eclipses occur between conjunction and opposition, if t is the actual interval between eclipses, T_1 the apparent interval between the eclipse at conjunction and the next eclipse at opposition, T_2 the apparent interval between that at opposition and that at the next conjunction, d the diameter of the earth's orbit, and V the velocity of light,

$$T_1 = nt + \frac{d}{V},\ T_2 = nt - \frac{d}{V},\ \text{and}\ T_1 - T_2 = \frac{2d}{V} = 33\text{m.}23.2\text{s.}$$

The best determinations of this so-called "equation of light" and of the

diameter of the earth's orbit give 298,300 kilometers per second as the velocity of light.

Bradley's Method. Römer's explanation was discredited until long after his death, when an entirely different astronomical method confirmed his views. In 1727 Bradley, the astronomer royal of England, discovered an apparent negative parallax of the fixed stars; that is, an apparent displacement not opposite to the direction in which the earth was moving in its orbit, but in the same direction. The apparent path of the stars in the ecliptic was back and forth in a straight line; of those near the poles of the ecliptic in circular orbits, and of those in intermediate positions in elliptical orbits. The displacement was the same for all stars in the same celestial latitude. These facts showed that the effect was not due to any proper motion of the stars, but rather to that of the earth in its orbit; and it is not a parallax effect, as the displacement is opposite to that of parallax. Bradley was for a time greatly perplexed by this phenomenon, but the chance observation of the direction of a wind vane on a boat sailing on the Thames, this direction not being that of the wind, but of the resultant of that of the actual wind and that of the virtual wind due to the motion of the boat, suggested to him that the apparent motion of the light coming from the stars might be the resultant of the actual motion of the light and its relative motion with respect to the moving earth. If a stone is dropped into a vertical tube which is at the same time moving parallel to itself in a direction at right angles to the path of the stone, the latter will have a horizontal component of relative motion with respect to the tube and will strike its side. Similarly a beam of light which actually moves with a finite velocity parallel to the axis of a telescope tube will strike to the side of the latter on account of its displacement due to the motion of the earth. If the apparent angular displacement is α, it is evident that $\tan\alpha = u/V$, where u is the component velocity of the earth at right angles to the line of sight and V the velocity of light. Bradley gave the name *aberration* to this apparent angular displacement of the light from the stars. The best determinations of α, the aberration constant, is 20.445″, which, combined with the known velocity of the earth in its orbit, gives a value for V of 299,920 kilometers per second.

410. Fizeau's Method. The first to make a direct determination of the velocity of light was Fizeau, who in 1849 found the time required for light to pass between Suresnes and Montmartre, near Paris, a distance of 8633 meters. His method was as follows: Light from a source S (Fig. 226) is reflected from a piece of plate glass m, focused by a lens L on the circumference F of a toothed wheel W, and, after passing between the teeth of the wheel, is made parallel by a second lens L_1. From this point the beam

travels to the distant lens L_2, which focuses it on a mirror M. From this point the beam retraces its path to the source; but a portion of it will pass through the plate glass m to the eye E, by which it may be observed. If the toothed wheel is rapidly rotated a detached train of light waves will pass through as an opening passes F, travel to M, and return. If in the meantime a tooth has moved

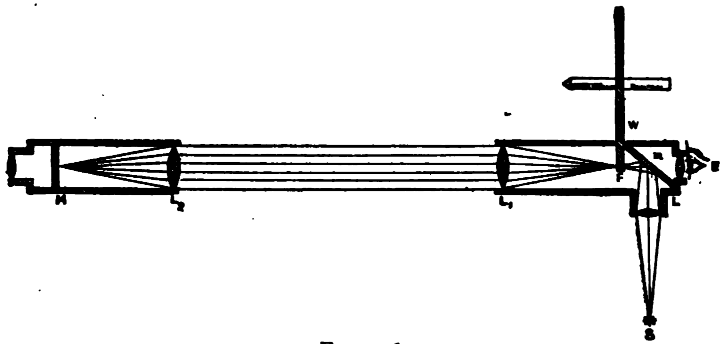


FIG. 226.

into the position F the light will be eclipsed; at twice the speed required for the first eclipse the light will again reach F when an opening is at the point, and will pass to the eyepiece. At three times the original speed of the wheel the second eclipse will occur, and so on. At speeds permitting transmission of the light the waves will pass and return through the successive openings in intermittent groups, but the light will appear continuous to the eye because of the persistence of vision. If L be the distance between the wheel and the distant mirror, V the velocity of light, and T the time required for light to travel the distance $2L$,

$$V = \frac{2L}{T}.$$

As the speed of the wheel increases successive eclipses occur. If n be the number of teeth and of spaces, and if N_1, N_2, and N_3 be the number of revolutions per second for the first, second, and third eclipse,

$$T = \frac{1}{2nN_1} = \frac{3}{2nN_2} = \frac{5}{2nN_3}.$$

$$V = \frac{4LnN_1}{1} = \frac{4LnN_2}{3} = \frac{4LnN_3}{5}$$

The value of V found by Fizeau was 313,300 km./sec. Cornu, using the same method, obtained a mean result of 299,950 km./sec. from several series of experiments.

Method of Foucault, Michelson, Newcomb. In 1862 Foucault determined V by means of the displacement of a beam of light reflected from a revolving mirror. The method was improved by Michelson, who made a series of observations in 1879 at the United States Naval Academy, and another in 1882 in Cleveland. Michelson's arrangement is indicated in Fig. 227. Light from a narrow slit S falls on the mirror m and is reflected to a lens L, which throws it in a parallel beam to the plane mirror M. The beam retraces its path, and if the mirror m is at rest is brought to a focus at S. If, however, m has rotated through the angle a while the light is passing from m to M and back, the reflected pencil will be rotated through the angle $2a$ and will form an image of the source at S_1. If the distance between S and $S_1 = d$, that between S and $m = r$, that between m and $M = L$, if n be the number of revolutions of m per second, and T the time required for light to pass from m to M and back,

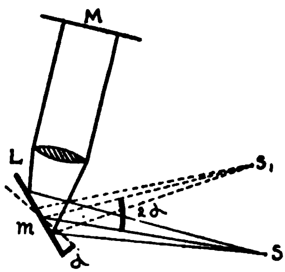


FIG. 227.

$$2a = \frac{d}{r}$$

$$T = \frac{a}{2\pi n}$$

$$V = \frac{2L}{T} = \frac{4\pi L n}{a} = \frac{8\pi L n r}{d}$$

Foucault used a short-focus lens between S and m instead of a long-focus lens between m and M, as in Michelson's arrangement; consequently L was a short distance, not exceeding 20 meters, and the displacement d was only 0.7 mm., even when the mirror revolved 800 times per second. The result obtained by Foucault was 298,000 kilometers per second. In Michelson's experiments a long-focus lens enabled him to make r large and

at the same time to throw a parallel or nearly parallel beam on M, so that the distance L could be increased indefinitely without any considerable loss of light. With a value of $L = 605$ meters, $r = 9$ m., and a speed of 257 revolutions per second, the displacement d was 133 mm. The result of Michelson's latest experiments in 1882 was $V = 299{,}850$ km./sec.

Newcomb, in 1882, made some further improvements in Foucault's method. The distance L was 3,721 meters, between the Washington monument and Fort Myer, in Virginia. The value of V obtained by him was 299,860 km./sec. The final results of Michelson and of Newcomb are probably not in error by more than 30 km./sec.

The method of Foucault is especially adapted to the measurement of the **Velocity of Light in Different Media,** such as water and carbon bisulphide, and was so applied by Foucault and Fizeau, and also by Michelson. A long tube filled with the liquid was placed between the mirrors m and M. Michelson found the velocity in air to be 1.33 times greater than that in water, and 1.76 times greater than that in carbon bisulphide. This has an important bearing upon the choice between the emission and the undulatory theories of light (§ 412).

The velocity of light from all sources seems to be the same, not being appreciably affected by their intensity. Römer and Bradley used sunlight or starlight, Fizeau and Cornu calcium light, Foucault, Michelson, and Newcomb sunlight, Young and Forbes electric light. In space lights of all colors travel with the same velocity. This is shown by the eclipse of a white star by the moon; the star would appear red just before eclipse and blue just after if blue light travels faster than red; but no change of color is observed. It is also shown by the fact that in Michelson's experiment the light was not drawn out in a spectrum. Photographs by Heyl of the spectrum of the variable star Algol, the light from which has a period of variation of about 69 hours, show that the intensities of the extreme violet and extreme red rise and fall simultaneously, proving that there is no relative retardation between them. In some material media the velocity of light of the different colors differs considerably. Michelson found the velocity of blue light in carbon bisulphide to be 1.4 per cent. less than that of red. In gases this difference is inappreciable.

Light reaches the earth from the moon in about one second and from the sun in about 8.25 minutes. A small parallax has been found in the

case of some of the nearer stars, which enables rough estimates of their distances to be made. Light from one of the nearest stars, α Centauri, would require about 3.75 years to reach the earth, and that from Sirius about 17 years. It seems quite possible that a distant star may have been destroyed by an explosion or collision generations ago, and yet be visible to us by light emitted before its destruction and still on its way through space. The changes frequently observed in variable stars must take place years before they are evident to us.

There is no reason to believe that there is an appreciable amount of matter of any kind between the earth's atmosphere and that of the sun or other heavenly bodies, nor is there any diminution, but rather a gain, in the amount of light transmitted through a glass vessel after exhaustion. It seems evident, therefore, that light will travel in what we call a vacuum.

THE NATURE OF LIGHT.

411. Mode of Transmission. According to some of the older hypotheses, such as that of Descartes, light is the effect of a pressure instantaneously transmitted through a universal medium. The fact that the disturbance producing light has a finite velocity shows, however, that it is due to motion, not to a static pressure. The radiation from such bodies as the sun heats substances on which it falls, and may produce chemical changes or electrical effects, which shows that a continuous stream of energy flows from luminous sources. According to our experience, there are only two ways in which energy may be transferred—by the actual projection of material bodies through space or by the transmission of vibrations or pulses through a stationary medium, as illustrated by different types of wave motion. Consequently there have been two rival theories regarding the propagation of light, the *emission* theory and the *undulatory* or *wave* theory.

412. Emission Theory. Sir Isaac Newton believed that light is due to the emission of luminous particles ("corpuscles") from the source. He appears to have adopted this hypothesis chiefly because it explained the rectilinear propagation of light, for which the wave theory seemed inadequate. Newton showed by prismatic analysis that white light is a combination of many different colors. He attributed difference of color to difference in size of the corpuscles exciting luminosity.

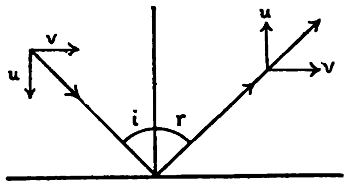


FIG. 228.

The emission theory satisfactorily explains reflection if we suppose the corpuscles to behave like elastic spheres. If such a sphere strikes a reflecting surface at an angle i with the normal (Fig. 228) the tangential component v of its velocity will not be changed. If the magnitude of the reflected component u is unaltered, it follows that the angle r of reflection is equal to the angle i of incidence.

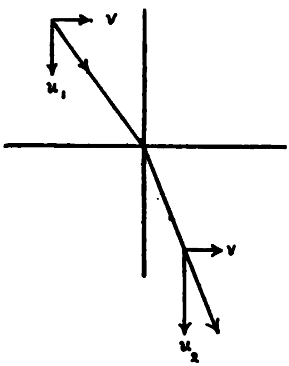


FIG. 229.

Refraction is also explained if we assume that matter attracts these particles. They will then be subject to a normal acceleration as they approach the boundary, while the tangential component of velocity is unchanged (Fig. 229). If the medium offers no resistance to the motion of the corpuscles (that is, if it is transparent) it follows that the increased velocity should be maintained after entering the second medium, and that the velocity of light should be greater in more refractive media than air than it is in the latter. The experiments of Foucault, Michelson, and others show that the opposite is true in all cases tested. This is one grave objection to the emission theory. Furthermore, if matter attracts light corpuscles, it would be difficult to account for the enormous expulsive forces required to project the particles from luminous sources. We should also expect the speed of the particles to vary with the nature and activity of the source; and yet the velocity of light from a candle appears to be the same as that from the sun.

413. Newton's Rings. In 1663 Robert Boyle described the brilliant colors observed in soap bubbles and other thin films, an effect which appeared to depend solely on the thickness of the films, not on their nature. Hooke made similar observations, which led him to believe that the effect is due in some way to an interaction between the light reflected from the upper surface of the film and that from the lower surface. About 1672 Newton investigated this phenomenon, which he tried, with poor success, to explain in terms of the emission theory. In order to secure a thin film of air, varying in thickness in a determinate manner from point to point, he pressed a convex glass lens of great radius

of curvature against a piece of plane glass. If light falls normally on such a combination, light of a given color is found to be reflected in a greater proportion than the other colors from all points where the film has a given thickness, the predominant color varying with the thickness. As the loci of points of equal thickness form circles about the region of contact, colored rings are observed concentric with this point. These have been called Newton's rings, or the colors of thin plates. Colored rings are likewise observed in the transmitted light. These are not so brilliant, however, as those due to the reflected light, as the transmitted colors are mixed with a large proportion of unmodified white light. The colors in the two sets of rings are complementary—that is to say, the light transmitted through a given point is white deprived of the color which is most strongly reflected from that point. If monochromatic light is used the rings are alternately dark and of the color used. In a wedge-shaped film these bands are parallel to the edge of the wedge; in a film of uniform thickness circular bands are produced under certain conditions, uniform color effects under others. These colors of thin plates are seen in all kinds of thin transparent films, such as soap bubbles, films of oil on water, and thin sheets of mica.

Newton attempted to explain this phenomenon by assuming that the light particles were put into "fits of easy reflection or of easy transmission" by the reaction on them of the waves set up in matter or in a surrounding ethereal medium by the impact of the particles on the surface. Such an explanation is manifestly very artificial, and could hardly have satisfied Newton himself; but to his mind there seems to have been a greater difficulty in accepting the wave theory, which had been more or less vaguely suggested by various persons, and first clearly formulated by Huyghens, a contemporary. Newton observed that water waves pass around obstacles without sensible disturbance, casting no shadows, and that sound shadows arise only under exceptional circumstances. Reasoning by analogy he could not see why light, if due to wave motion, should not travel around corners instead of in straight lines. He noticed, however, that sound waves had a greater tendency than water waves to cast shadows, and if he had carefully observed the behavior of small waves, such as ripples on water, his objections to the wave theory would probably have been removed. While large water waves pass around a pile or other comparatively small obstacle, ripples are effectually stopped, passing the object on each side without reuniting; there is a well-defined region of no disturbance, or shadow. Similarly, sounds of high pitch, due to very short waves, cast well-defined shadows.

Again and again Newton seems to have been attracted by the undulatory theory, but his final acceptance of it was prevented by the fact of rectilinear propagation.

414. Wave Theory of Light. Huyghens distinctly formulated this theory about 1678. He believed that space is filled with a rare medium, the ether, through which the waves are propagated from luminous bodies. This theory accounts without any difficulty for the ordinary phenomena of reflection and refraction, but was not acceptable to Newton for the reason above stated. Huyghens' ideas will be further elaborated in discussing the work of Young and Fresnel.

415. Interference. For more than a century after Newton's time little progress was made in the subject of light, until, in 1802, Thomas Young published a paper "On the Theory of Light and Colors" in the Philosophical Transactions of the Royal Society of London. In this he discussed optical phenomena from the standpoint of the wave theory, and first called attention to the fact, overlooked by Huyghens and other advocates of the wave theory, that the effect at any point of space through which light waves are passing is the resultant of the effects of a number of coincident individual waves. The magnitude of this resultant depends not only on the amplitudes, but also on the relative phases of the component waves. If two waves of equal amplitude and moving in the same direction are in the same phase the displacement at any point is the sum of the individual displacements, and the energy, which is proportional to the square of the amplitude, is four times as great as in a single wave. If the waves are opposite in phase, the resultant amplitude and energy at any point are zero. This effect Young called the *interference* of light waves.

He showed that the colors of thin plates can be very simply explained by taking this into account. Consider for example a thin air film of thickness t, on which light falls normally or nearly so. A portion of the light will be reflected at the first surface along AB (Fig. 230) and another portion from the second boundary along CDE. If the surfaces of the film are parallel these two components will emerge parallel to each other and will be brought to a focus at the same point of the retina of an eye E focused for infinity. If the geometrical difference of path, which is evidently $2t$ for normal incidence, is any whole number of wave-lengths of a particular

color, and if there is no retardation in the act of reflection of either component, that color will be reënforced in the reflected light, being about four times as intense as in the case of ordinary reflection from one surface, if these two components alone be considered; but, as shown in the figure, there are a number of other components *FGH*, etc., which have undergone three, five, etc., internal reflections, and which emerge in the same direction and reënforce the two components first considered; so that there is a more than fourfold reënforcement of that color. If, however, the difference of path between the reflected portions be an odd number of half wave-lengths, the corresponding color will be subject to destructive interference and will be greatly weakened in the reflected light. For the transmitted components *CH*, *FI*, etc., it is seen that the geometrical difference of path is also equal

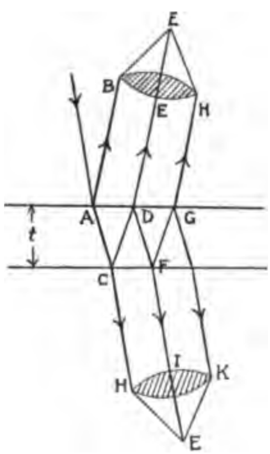


FIG. 230.

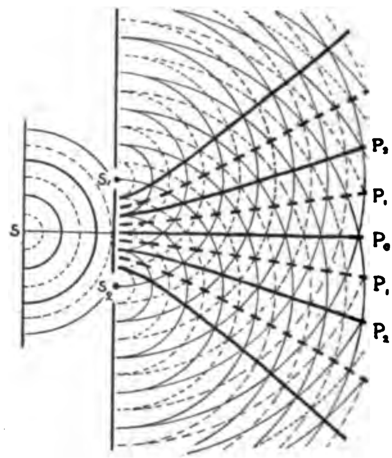


FIG. 231.

to $2t$ for normal incidence, so that we should expect for this reason a reënforcement of the transmitted light likewise—a result which would hardly be in conformity with the law of conservation of energy. Observation shows that the contrary is the case, the reënforcement of a color reflected from a given point always corresponding to a weakening of the color transmitted at that point. Young showed that this remarkable fact is due to a loss of time of one-half period in one of the reflections. For a more complete discussion of the colors of thin films see § 470.

416. Young's Experiment. Young also devised a simple experiment which may be regarded as a crucial test of the wave theory. Light diverging from the slit S (Fig. 231), which acts as a primary source, passes through two narrow slits S_1 and S_2 very

close together, which act as secondary sources. If a screen be placed beyond these slits a series of colored and dark bands parallel to the slits will be observed on it. If one of the slits is covered the bands disappear. This shows that they are the resultant effect of two superimposed pencils of light alternately reënforcing and destructively interfering with each other.

It is easy to repeat Young's experiment by ruling two narrow slits very close together on a developed photographic plate and looking through these slits at a distant electric light. The explanation is as follows: Through the slits S, S_1, and S_2 the wave disturbance propagates itself in all directions beyond the respective screens in semi-cylindrical waves having these slits as axes. That this is the case may be seen by holding a white screen in front of such a narrow slit on which light falls. It will be seen that the transmitted light does not pass in a narrow pencil, but actually diverges very considerably from the axis of the pencil, the amount of divergence increasing as the slit is narrowed (§ 479). Similarly it may be shown that if an extended water wave or ripple in a mercury surface strikes a screen with a small opening in it circular waves will diverge from this opening as a center (Fig. 199) the most intense effect, however, occurring along the axis SP_0. There are, consequently, when two slits are used, two sets of semi-cylindrical light waves diverging from these slits and crossing each other, as shown in Fig. 231. Along SP_0 every point of which is equidistant from S_1 and S_2, waves of all lengths from the two sources will always meet in the same phase, and there will be a maximum of white light on the screen at P_0. Along the dotted line ending at P_1 the distances of any given point from the sources differ by half a wave-length; there is destructive interference along this line and a minimum for the corresponding color at P_1. Along the line ending at P_2 the difference between the distances of any given point from the sources is a whole wave-length, so that along this line waves of the same length meet in the same phase and there is a maximum for the corresponding color at P_2. At any point P_n for which $S_1P_n - S_2P_n = n\lambda$ (n being any whole number) there will be a maximum; where $S_1P_n - S_2P_n = \frac{1}{2}(2n + 1)\lambda$ there will be a minimum. It is evident that these lines of maximum and minimum effect are hyperbolas having S_1 and S_2 as foci. It must be noted that this effect is dependent on a succession of waves through S_1 and S_2. The first wave from S through S_1 reaches P_1 at the same instant that the second wave from S through S_2 reaches that point. If only one wave were to come from S the resultant single waves from S_1 and S_2 would reach P at different instants and could not interfere.

If $x = P_0P_2$ is the distance between maxima of a given wave-length,

$L_0 = AP_0$ the distance of the screen from the slits, $b = S_1S_2$ the distance between the slits (Fig. 232), and if we assume the angle at D to be a

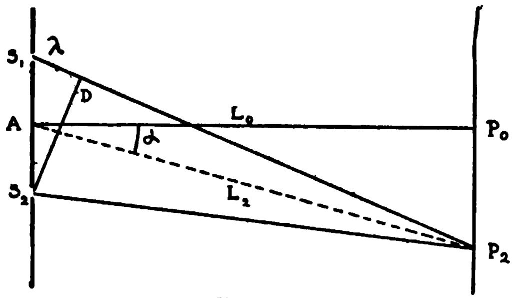


FIG. 232.

right angle (this is very nearly true if b is very small) we have very nearly

$$b : \lambda = L_2 : x = L_0 : x$$

since, because the angle α is very small, L_0 is very nearly equal to L_2. Therefore,

$$x = \frac{L_0}{b}\lambda.$$

This expression gives the width of a band of wave-length λ. For a given screen distance this width is directly proportional to the wave-length and inversely proportional to the distance between the slits.

These results may be illustrated by other kinds of waves; for example, mercury ripples (§ 361). A similar experiment may be performed with sound waves set up by a whistle of high pitch and passing through two openings. Points of maximum and minimum disturbance may be found beyond the screen by means of a sensitive flame.

417. Relation between Color, Wave Length, and Frequency. If white light falls on the slits the inner side of each band is violet, the outer side red. This shows that the wave length is different for light of different colors, and that the wave length of violet light is less than that of red. The central band is of course white, as all colors have a maximum at this point, regardless of their wave length. From the relation $n\lambda = V$ (§ 349) where n is the frequency of vibration, λ the wave length, and V the velocity of light, it is evident that when V changes either n or λ or both must change. If Young's experiment be performed in a medium such as water, it is found that the width of the bands in water is to their width in air as the velocity of light in water is to that in air. Hence $\lambda_1/\lambda = V_1/V$, and n is constant. It is a matter of common ex-

perience that the color of a beam of light does not change when it enters water, hence frequency rather than wave length determines color. Color is, therefore, analogous to pitch in sound.

418. Interference Effects. Fresnel, a young French artillery officer, about 1815, produced these effects by light diverging from a slit S and reflected from two adjacent mirrors MM inclined at such a great angle as to be almost in the same plane. As shown by Fig. 233, the light arriving at any point P where the pencils overlap appears to come from two virtual sources S_1 and S_2, the effect of which is precisely the same as that of the two real sources in Young's experiment. (The term *virtual* source applies to a point from which the waves appear to diverge without really originating at that point.)

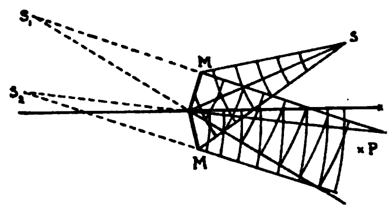


FIG. 233.

Fresnel also produced interference effects by the use of a biprism B, equivalent to two prisms of very small refracting angle placed base to base (Fig. 234). Here, again, it is evident that the transmitted light appears to come from the two virtual sources S_1 and S_2.

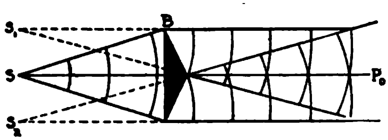


FIG. 234.

Another method of obtaining similar interference effects is by means of a convex lens L cut along a diameter in two halves which are slightly separated, giving two real or virtual images of the source, from which the waves diverge and overlap. This is known as the Billet split lens (Fig. 235).

The interference effects due to Lloyd's single mirror (Fig. 236), are caused by waves coming respectively from the real source S and the virtual source S_1. The fringes are easily obtained by reflecting the light from a narrow slit or lamp filament at grazing incidence from a mirror of black glass, in order that the effects may not be complicated by reflection from the rear surface.

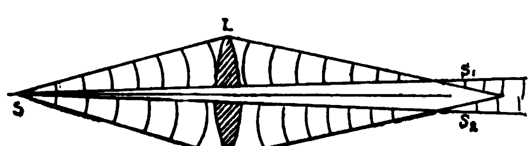


FIG. 235.

These various methods may be used for rough determinations of the wave-lengths of the different colors, as was first done by Young. It is thus found that these waves are very small, the average wave-length of red light being about 0.000065 cm., and that of the violet about 0.000042 cm.

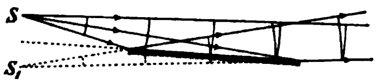


FIG. 236.

419. The Ether. To account for the transmission of waves through space containing no ordinary matter it seems necessary to assume the existence of a universal medium filling all space and even interpenetrating matter itself, as shown by the existence of transparent substances. That this medium can react on matter is shown by the fact that radiant energy is transmitted from ether to matter in the case of absorption, and from matter to ether in the case of emission of radiation by material sources. This medium appears to be like a jelly in the respect that it will transmit transverse but not longitudinal waves. We can infer its properties only from optical and electrical phenomena, as it is not tangible or visible; but it may be questioned whether, after all, the effect of light waves as they strike the retina is not as direct evidence of the existence of the ether as a blow with a club is evidence of the existence of matter.

420. Huyghens' Principle. Huyghens assumed that a wave is propagated by every point of the medium in a wave front acting as a new center of disturbance (§ 358). The resulting wave front is the enveloping tangent plane to the wavelets starting from these centers, as shown in Fig. 237. While Huyghens evidently considered the wave front as a resultant effect due to all these contributing wavelets, he does not seem to have had a clear idea of the destructive interference between wavelets at points where they meet in different phases. This conception seems to have first occurred to Young, and later but independently to Fresnel.

Huyghens' principle is easily illustrated experimentally, as shown in Fig. 199. The only element of the wave incident on the screen which has an opportunity to affect the medium beyond the latter is that which is incident at the opening *S*, so that beyond the screen the undisturbed effect of this element may be observed. In such a case a series of semi-circular waves diverges from the opening on the surface of a liquid, or a series of hemispherical waves in the case of sound or light disturbances passing

through a small circular opening. The amplitude of these waves is greatest along the normal to the screen, gradually diminishing on each side of the axis.

By the application of Huyghens' principle it is easy to show why waves passing through an opening small in comparison with the wave-length diverge in all directions, while waves passing through an opening large in comparison with the wave-length travel in a straight "beam," as in the case of the so-called rectilinear propagation of light.

421. Extended Wave Front. The points *a, b, c,* etc., between *A* and *B* (Fig. 237) taken as close together as we please, act as centers of disturbance. Along the tangent plane *A′B′* the different waves are all in the same phase, and each point in this new tangent plane becomes a new center of disturbance, so that the resultant wave travels forward as rapidly as the disturbance is propagated from point to point of the medium. The waves move forward without hindrance, because there is no existing displacement to oppose them; they do not travel backward, because there is a force due to the existing displacement on the side from which the waves come sufficient to nullify the backward component of the displacement due to each successive center of disturbance. It is like the propagation of a shove through a line of people, or of elastic spheres of the same mass and elasticity; that in front is not braced to withstand the impulse, while the reaction on the one communicating the impact is expended in overcoming its forward momentum.

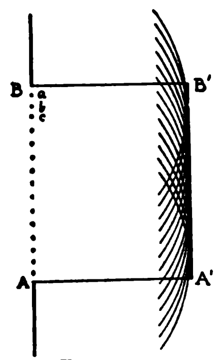


FIG. 237.

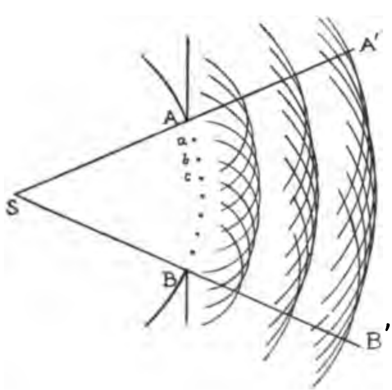


FIG. 238.

422. Waves through Large Opening. **The points *a, b, c,* etc., in the wave front *AB* (Fig. 238) act as centers of disturbance and propagate wavelets to**

the tangent surface $A'B'$. It is evident from the figure that only those wavelets between the lines AA' and BB' conspire in a common wave front. Outside of these lines the wavelets cross each other in all directions and in all possible phases at random, so that the resultant disturbance is zero except in the immediate neighborhood of AA' and BB', where certain diffraction effects are produced which will be discussed later (§ 476).

423. Waves through Small Opening. From Fig. 239 it is clear that no opposition of phase between the elementary wavelets from *a*, *b*, *c*, etc., can arise until the point P_1 is reached, where the difference between AP_1 and BP_1 is half a wave-length, and even then the disturbances from the extreme points *A* and *B* alone are in opposite phase. Only when this difference of path is a whole wave-length can complete destruction arise. In this case we see that the disturbances from the two halves of the opening reach P_1 with an average difference of path of half a wave-length, so that the wavelets nullify each other pair by pair. If the slit is less than a wave-length in width some effect is produced even at the point P_2. The effect is evidently always greatest at P_0, where the wavelets meet very nearly in the same phase, and least at P_2 where there is the greatest diversity of phases (see § 479).

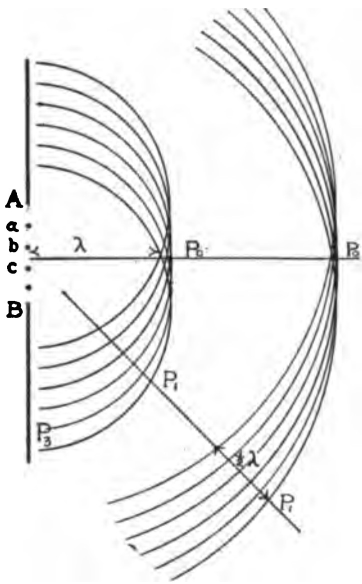


FIG. 239.

424. Polarization. In the case of sound waves the displacement of the vibrating medium is longitudinal, or in the direction of the propagation of the wave. In a stretched wire the vibrations may be longitudinal or transverse. In waves in liquids there is a combination of longitudinal and transverse displacements, as illustrated by the rotary motion of particles in water waves. Young, having in mind the analogy with sound, seems to have tacitly assumed at first that the displacement of ether in light waves is longitudinal. While this assumption involves no contradiction with experimental results when applied to reflection, refraction, and interference, it fails as completely as the emission theory in the explanation of an effect known as *double refraction* discovered by Bartholinus in 1669 and later investigated in more detail by Huyghens (see § 530). Newton concluded from these phenomena that light is so modified by passing through a doubly-refracting crystal that its properties are not the same in all

azimuths in a plane perpendicular to the direction of propagation. Newton endeavored to explain these effects by attributing a sort of polarity to the light corpuscles; consequently light thus modified by passage through a double-refracting crystal is said to be polarized.

425. Direction of Vibration. From the standpoint of the wave theory it is evident that longitudinal vibrations, such as those of sound, bear the same relation to the medium through which they are propagated in all azimuths in a plane perpendicular to the direction of motion, so that they could not possibly be affected by the rotation of the medium in that plane. If sound waves pass through a fence with pickets vertical they will pass with equal facility through a fence with pickets horizontal. This is not true with respect to transverse vibrations, which take place in a definite direction at right angles to the line of propagation. If a cord passes between the vertical pickets of a fence transverse waves in it will freely pass through the fence if they are vertical, but will be stopped if they are in a horizontal plane. From such considerations applied to the phenomenon of double refraction and that of polarization (§ 530) Young and later Fesnel concluded that the vibrations of light waves must be *transverse*. This conception removed all difficulty from the explanation of polarized light.

426. Sources of Light are usually bodies at high temperatures. It is believed that high temperature is the effect of a violent agitation or vibration of an ultimate particle or molecule of matter either as a whole or as an integral part. We may imagine that these vibrating particles impart their motion to the surrounding ether in much the same way that a vibrating tuning fork generates sound vibrations in the air.

From the evidence now before us it seems reasonable to assume as a working hypothesis that light is due to a transverse wave motion in a universal medium; that the different colors correspond to different wave-lengths (or, more properly, to different rates of vibration); that waves of all lengths travel with the same velocity in free space, but with different velocities in matter; that these waves are excited by the vibrations of the particles of material sources, these particles being molecules or atoms (or in some cases even smaller bodies) (§ 550); that these particles in general vibrate in different planes and directions, and that the vibrations of a given particle may constantly change in direction; that each vibrating particle sends out into the surrounding medium a series of waves vibrating in the same plane as the particle, so that ordinary white light consists of a mixture of waves of many lengths, the resultant vibrations being in a plane at right angles to the direction of propagation, successive trains of waves having different planes of vibration; and that by double refraction or reflection we may sift out component vibrations in a given plane, and produce what is called polarized light. It will be found that all the experimental facts to be considered later are in harmony with these assumptions.

We shall next consider reflection and refraction from the standpoint of the wave theory. It will be found that these phenomena can be explained in a much clearer manner by this method than by that of "geometrical optics" in which it is assumed that light travels in straight lines called "rays"— an idea more consistent with the emission than with the wave theory. In the following pages the word *ray* will often be used as a matter of convenience, *meaning thereby merely a normal to the wave front, which indicates the direction in which the wave is moving at the point considered.* This definition applies only to isotropic media (§§ 163, 538).

REFLECTION.

427. Reflection from a Plane Surface. A wave diverging from the source S (Fig. 240) falls on a plane mirror MN. If the mirror

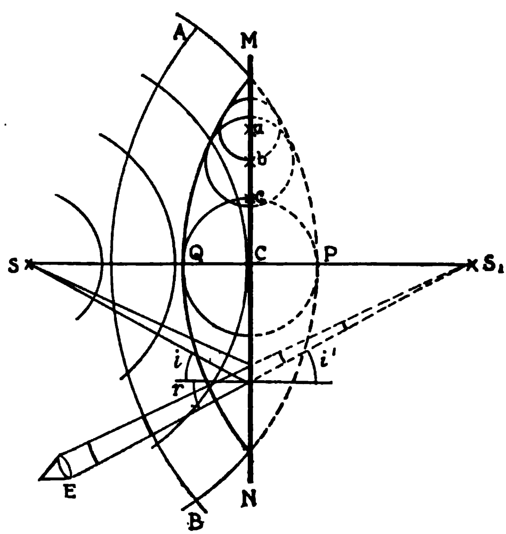


FIG. 240.

were absent, the wave would at a given instant occupy the position $AMPNB$. With the mirror in place, each element of the original wave when it reaches the mirror becomes the center of a reflected wavelet, just as it would have contributed a wavelet from the same point to form the resultant wave MPN if the mirror

were absent. If, therefore, a number of circles tangent to MPN be described about centers a, b, c, etc., on MCN they must touch both the imaginary wave MPN and the reflected wave MQN. The arcs MPN and MQN are evidently similar and equal and have equal radii of curvature. If S_1 is the center of curvature of the reflected wave $SC = CS_1$, the line SS_1 is normal to the mirror, and S_1 is as far behind the latter as S is in front of it. If the eye is at E, any point reached by the reflected wave, the pencil of light entering the pupil will be focused on the retina. As the vertex of this cone is virtually at S_1, the image of the source will appear to be at that point. From the diagram it is evident that the angles i and r are equal.

428. Focus. The source or center of curvature of a family of waves, either divergent or convergent, is called a *focus* —literally a hearth or source of radiation. The point S from which the waves actually come is called a *real focus;* the point S_1 from which they appear to come is called a *virtual focus.* The points S and S_1 are *conjugate foci.* Since the conjugate focal distances in the case of a plane mirror are equal, it is evident that if the mirror be displaced a given distance parallel to itself the image will be displaced twice that distance.

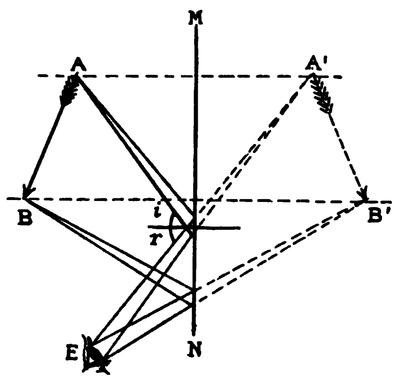


FIG. 241.

429. Images. If $A'B'$ is the image of AB (Fig. 241), it may be shown as above that the image of each point is as far behind the mirror as the point itself is in front, and on the same normal; and that, consequently, the image and the object are symmetrically placed with respect to the mirror and of the same size.

430. Multiple Reflection. Fig. 242 shows how these images are situated in the case of multiple internal reflection from surfaces AB and CD parallel to each other. The position of these images is readily determined by the fact that the image of the first order

in each surface is as far behind the surface as the source is in front, and on the same normal to the surface. The two images of the second order are fixed in the same way, by considering the images of the first order to be the sources, and so on *ad infinitum*. It is easy to see that when the mirrors are inclined at an angle a (Fig. 243) there are multiple images of the mirrors as indicated

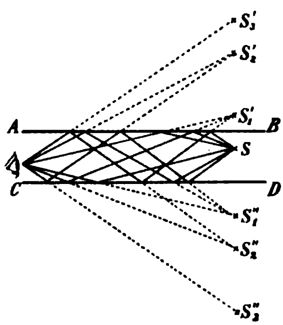


FIG. 242.

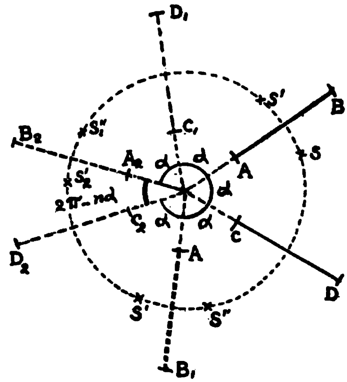


FIG. 243.

by the dotted lines, and that the successive images are symmetrically placed on each side of each mirror image and located in a circle about the point of intersection of the mirrors.

431. Reflection from Curved Surfaces. If a wave is reflected from a curved surface the curvature of the reflected wave is changed, unless it exactly conforms to the mirror surface at incidence. Experience shows that only in a few cases is the reflected wave spherical or approximately so, and only in such cases can a definite image be formed. The ordinary type of curved mirror is that with a spherical surface. The reflected waves are approximately spherical if the diameter of the mirror is small compared with its radius of curvature. In order to determine the position of the center of curvature and the conjugate focal relations for spherical mirrors a very simple mathematical relation is all that is required.

432. Relation between Radius of Curvature and Sagitta of Arc. Consider the arc AB, with center of curvature C, and radius r (Fig. 244). The distance x on the bisecting radius of the arc

included between the arc and the chord $AB = 2y$ is called the sagitta of the arc. To determine the relation between r and x write

$$r^2 = y^2 + (r - x)^2 = y^2 + r^2 - 2rx + x^2$$

Therefore,

$$2rx - x^2 = y^2$$

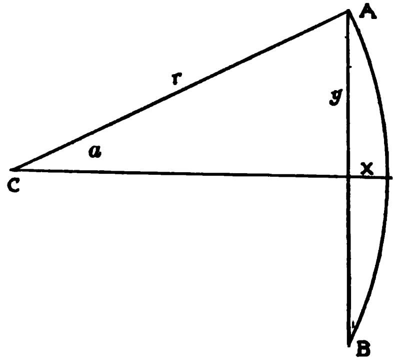


FIG. 244.

$$x = \frac{y^2}{2r - x} = \frac{y^2}{r(1 + \cos a)}$$

It is found that if the angle a is very small, not more than two or three degrees, the mirror will give a well-defined image. If the angular aperture $2a$ of the mirror is greater than four or five degrees spherical aberration becomes noticeable (§ 438). For all mirrors which give satisfactory images x may be neglected in comparison with r, or $\cos a$ regarded as equal to unity, so that within the limits of errors of measurement

$$x = \frac{y^2}{2r}$$

433. Concave Mirror. The source is at a distance u from a concave mirror MN (Fig. 245) with center of curvature at C and radius r. The waves incident on the mirror have a radius of curvature u, with a sagitta AB. Reflection begins at M and N while the vertex of the wave has still to travel the distance BD before reflection begins. When the vertex reaches the mirror the edges of the wave have travelled a distance $BD = AD - AB$ along MS_1 and NS_1. If the reflected wave is spherical it must have a definite center of curvature S_1 and radius v, with sagitta DE. At

the instant when reflection begins at D the incident wave, the mirror, and the reflected wave have a common point of tangency at D. If the angular aperture of the mirror is so small that the cosines of the angles α, β, and γ may be considered as equal to unity (these angles are exaggerated in the figure for the sake of clearness) we may consider the portions of the wave reflected from

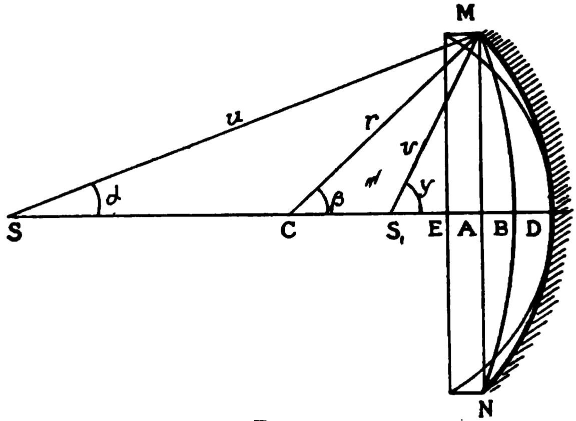


FIG. 245.

M and N to move parallel to the axis rather than in the directions MS_1 and NS_1. Hence

$$AD - AB = DE - AD$$
$$AB + DE = 2AD$$

It is not convenient to measure sagittæ, but by using the relation developed in § 432 the above expression can be transformed into one involving only the easily measured distances r, the radius of curvature of the mirror, u, that of the incident wave, and v, that of the reflected wave. The semi-chord y has the same value for all the arcs concerned, so that the common factor $y^2/2$ may be cancelled when $y^2/2r$ is substituted for AD, with similar substitutions for BD and DE. The final result is

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}$$

The justification for the somewhat inexact assumptions made in deriving this formula is found in the fact that it agrees with experimental observations within the limits of error of measurement.

A beam of light is always reversible in direction, hence, if the source is at S_1, the image will be at S.

If the source is at a great distance from the mirror the incident wave is practically plane (parallel beam), and u is infinite. The corresponding value of v is called the ***principal focal distance*** f. S_1 is then the ***principal focus***. The above equation then becomes

$$\frac{1}{\infty}+\frac{1}{f}=\frac{2}{r} \quad \text{or} \quad f=\frac{r}{2}$$

hence the principal focal point is half way between the mirror and its center of curvature. The conjugate focal relation may now be written

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$

If $u>r, v<r$.
If $u=r, v=r$.
If $u<r, v>r$.
If $u=f, v=\infty$.
If $u<f$, v is a negative quantity.

Fig. 246, which illustrates the last case, shows that the center

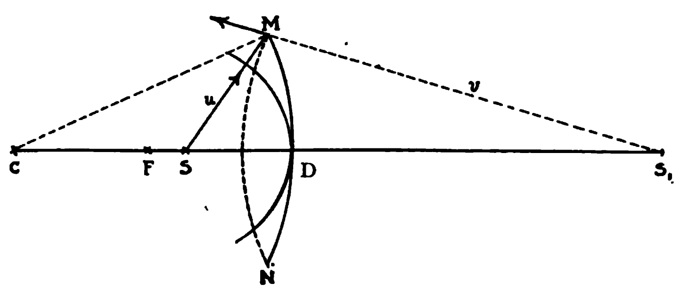


FIG. 246.

of curvature of the reflected wave is behind the mirror. It is a virtual focus, since the waves do not actually diverge from that point. It is clear that the negative sign of v indicates this result, since the distance $DS_1=v$ is measured in a direction opposite to that in which light actually proceeds after reflection, so that the reflected light cannot pass through the point S_1.

Writers differ in their conventions regarding the signs of conjugate focal distances and radii of curvature. The most easily

remembered and applied as well as the most consistent rules seem to be the following:

(*a*) *Distances measured from the source to the mirror and from the mirror in the direction toward which the light is reflected are considered positive.*

(*b*) *The radius of curvature of a converging surface is considered positive, that of a diverging surface negative.*

From the first rule it is evident that a positive value of v or f indicates a real focus, a negative value a virtual focus.

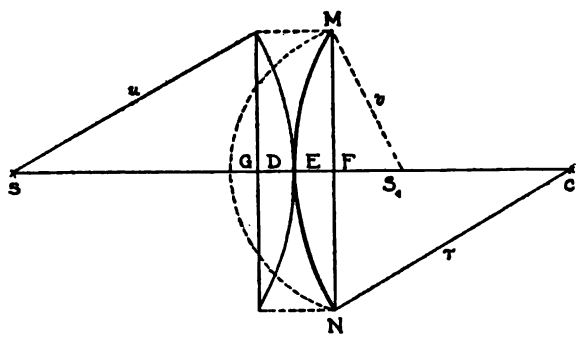


FIG. 247.

The second rule is justified by considering the case of a convex mirror, which, as shown by Fig. 247, has a divergent effect on the incident light.

434. Convex Mirror. Proceeding as in the previous case, if FG is the sagitta of the reflected wave and v its radius,

$$DE + EF = FG - EF$$

$$DE - FG = -2EF$$

$$\frac{1}{u} - \frac{1}{v} = -\frac{2}{r} = -\frac{1}{f}$$

where, provisionally, v, r, and f, may be considered as mere magnitudes affected with the negative signs in the formula.

Comparing this expression with that deduced for a concave mirror, we see that it will become identically the same if we agree to consider the radius of curvature and the principal focal distance of a convex mirror as negative; v is also negative since the reflected light is divergent.

The general formula applicable to all mirrors is, therefore,

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$$

u being always essentially positive, and f to be taken as positive for a concave, negative for a convex mirror. When $f=\infty$ we have the case of a plane mirror.

If we make f negative in the expression for v given in § 433, we see that in the case of a convex mirror v is always less than f and negative. If, however, the light incident on the mirror is convergent to a point at a distance $-u$ behind the mirror, v may become positive; so that a convex mirror may give a real image of a virtual source.

435. Geometrical Method. The same results may be obtained by applying the law of plane reflection to "rays" without any hypothesis as to the nature of light. The ray SD (Fig. 248) will, as it is incident normally at

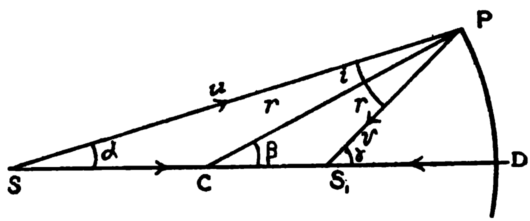


FIG. 248.

D, be reflected back on itself. The ray SP will be reflected at P, so that the angles i and r are equal. The intersection of these two reflected rays will fix the position of the image S_1. From a well-known geometrical relation we have

$$\frac{SC}{\sin i}=\frac{SD-CD}{\sin i}=\frac{u-r}{\sin i}=\frac{u}{\sin\beta}$$

$$\frac{CS_1}{\sin r}=\frac{CD-S_1D}{\sin r}=\frac{r-v}{\sin r}=\frac{v}{\sin\beta}$$

Therefore,

$$\frac{u-r}{r-v}=\frac{u}{v}$$

From which

$$ur+vr=2uv$$

Dividing through by uvr,

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{r}$$

The assumptions made in this case are that $SD = u$ and $S_1D = v$, which is a sufficient approximation to the truth when the angles α, β, and γ are small.

The formula for the conjugate focal relations of a convex mirror may be derived in the same way.

In some cases it is more convenient to use the geometrical or ray method than that of waves; but it must always be remembered that these "rays" merely represent normals to the wave front.

436. Images Formed by Spherical Mirrors. If any two radii be drawn from any point of a source, the point of their intersection after reflection will fix the position of the corresponding point of the image. Any pair of radii will do, but for convenience two of the following are usually chosen, because their course after reflection is easily determined: The radius parallel to the axis, which after reflection passes through the principal focus; that which passes through the principal focus, which becomes parallel to the axis after reflection; that which is incident to the intersection of the mirror with its axis.

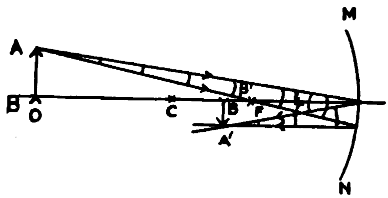


FIG. 249.

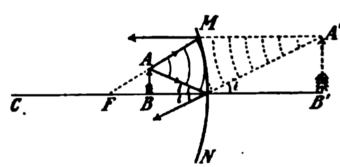


FIG. 250.

The construction of the images formed by a concave and by a convex mirror is illustrated by Figs. 249, 250, 251, where the points A' and B' are located by using the pair of rays last mentioned above. In the first, the image is real and inverted; in the second and third the images are virtual and erect.

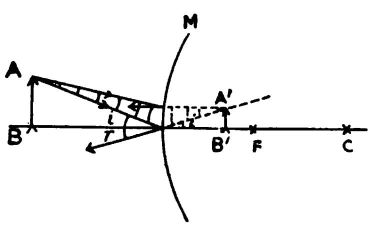


FIG. 251.

437. Magnification. Since the angle subtended by the object at the mirror is i, while that subtended by the image is the equal angle

r, it is evident that the relative sizes of the object and image are to each other as their respective distances from the mirror.

$$o/i = u/v$$

The real image formed by a concave mirror may be of the same size of the object, or larger, or smaller; the virtual image is always larger, since $v > u$. The virtual image formed by a convex mirror is always smaller than the object.

438. Spherical Aberration and Caustic Curves. If a converging wave is truly spherical there is a perfect focus at its center of curvature. As a matter of fact, the waves reflected from a spherical mirror are not perfectly spherical, except in the special case where the source is at the center of curvature of the mirror. Consider, for example, a plane wave falling on a concave mirror MN of radius r (Fig. 252). If the ray PO parallel to the axis is incident at the point O the reflected ray will intersect the axis at F_1 and the angle of incidence i and that of reflection r will be equal to the angle α between the radius CO and the axis. To find the distance of the point F_1 from the mirror write

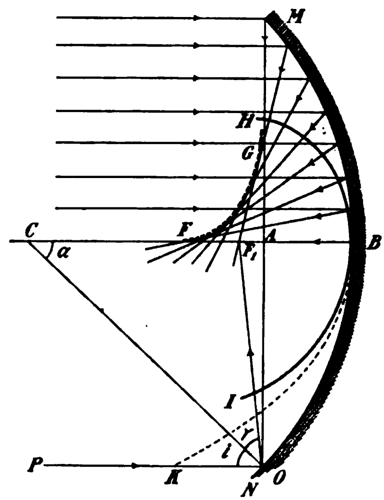


FIG. 252.

$$CF_1 \cos \alpha + OF_1 \cos r = 2CF_1 \cos \alpha = r$$

or

$$CF_1 = \frac{r}{2 \cos \alpha}$$

$$\therefore \; F_1B = r - CF_1 = r\left(1 - \frac{1}{2 \cos \alpha}\right) = r\,\frac{2 \cos \alpha - 1}{2 \cos \alpha}$$

When α is very small $F_1B = FB = r/2 = f$. As α increases F_1B diminishes. Hence rays incident at a point distant from the axis will intersect the latter at a point between the principal focus F and the mirror, as shown in the figure. The surface FG enveloping the bundle of inter-

secting rays is called a *caustic* surface, and the cross-section, with cusp at the principal focus F, is called a caustic curve. The actual reflected wave surface is everywhere normal to these rays, and is indicated by the full line BH in the upper half of Fig. 252. The section of the hypothetical spherical surface BI concentric with F is shown by the full line in the lower half of the figure. The section of the surface corresponding to the relation $KO = AB$ from which the conjugate focal relations were derived is shown by the dotted line. All these curves are practically coincident near the axis, so that a definite focus is obtained when the angular aperture $2a$ of the mirror is small.

The deviation from spherical shape of waves reflected from a mirror of large aperture is called *spherical aberration*. The effect in case of a concave mirror is evidently greatest when the incident wave is plane (excluding the unusual case of an incident convergent wave), and decreases as the distance of the source from the mirror decreases, vanishing when it reaches the center of curvature.

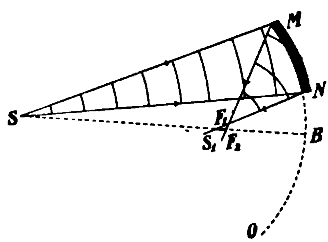


FIG. 253.

The light reflected from the sides of a cup containing coffee or milk plainly shows the caustic curve on the surface of the liquid.

439. Focal Lines. When waves from a point source are incident obliquely on a spherical mirror, it is found that no point image is formed, but that there are two line images in different positions and at right angles to each other. These are called *focal lines*. The origin of these lines is clearly seen if we consider the mirror MN (Fig. 253) to be part of a larger mirror MNO, on the axis SB of which the source S lies. Constructing the reflected rays incident at different points of this mirror, it is clear that, while the focal cusp of the entire mirror is at S_1, all the rays coming from MN intersect approximately at the point F_1. The diagram gives a cross-section of the incident and reflected rays. If this diagram be rotated about the axis SB by an amount equal to the diameter of the mirror MN the point F_1 will describe the arc of a circle with its center on the line SB. This is the primary focal line, which will appear on a screen placed at F_1 as a narrow curved strip. After passing F_1 all the rays reflected from MN will intersect the axis SB at various points between S_1 and F_2 (since all the planes of incidence contain SB). A screen placed at this point will show a narrow elongated patch of light, S_1F_2, the secondary focal line. If the screen is at right angles to the reflected pencil the patch of light will be approximately a lemniscate or figure 8.

440. Cylindrical Mirror. A parallel beam incident on such a

surface is brought to a real or virtual line focus. The image of a point source is likewise a line. Such mirrors and the reflected pencil are said to be *astigmatic*. (A pencil symmetrical about an axis, that is, having a point vertex, and thus giving a point image of a point, is said to be *homocentric*.) In the case of a concave cylindrical mirror, if the point source lies outside the principal focus, there will be a real image AB and a virtual image $A'B'$ in planes at right angles to each other, as illustrated in Fig. 254.

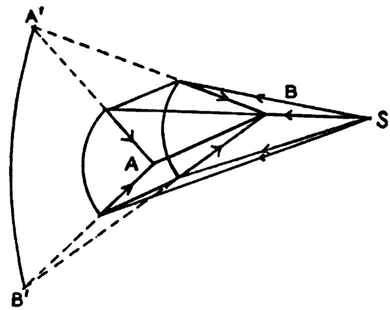


FIG. 254.

441. Paraboloidal, Ellipsoidal, and Hyperboloidal Mirrors. The light from a point source at one focus of an ellipsoidal reflector will be brought without aberration to the other focus, a real image being formed. Light from a source at one focus of a hyperboloidal mirror will have a virtual focus at the conjugate focus of the mirror. If the source is at the focus of a paraboloidal mirror, the light will be reflected in a parallel beam; and parallel light will be brought without aberration to a real focus by such a mirror.

REFRACTION AND DISPERSION.

When light waves fall on material bodies they are in part reflected and in part absorbed or transmitted. When light passes through transparent bodies it is supposed that the vibrations of the light waves are transmitted through ether interpenetrating the matter. It appears, however, that the velocity of the waves is affected by the presence of matter, as shown below, and that in all cases more or less of the vibratory energy of the ether is communicated to the particles of matter. This constitutes *absorption*.

442. The ancients were acquainted with the fact that a beam of light is more or less deviated in passing from air to water. The **Law of Refraction** was first discovered in 1621 by *Willebrod Snell*. He found by experiment that the ratio of the sines of the angles of incidence and of refraction is constant at the boundary between

two media. The ratio $\sin i/\sin r$ is called n, the *index of refraction.* The angle of incidence is usually measured in air.

It was shown by Huyghens that refraction is very simply explained by assuming a change of velocity in passing from one medium to another. Direct measurements by Foucault, Fizeau, and Michelson show that light travels with different velocities in air, water, and carbon bisulphide.

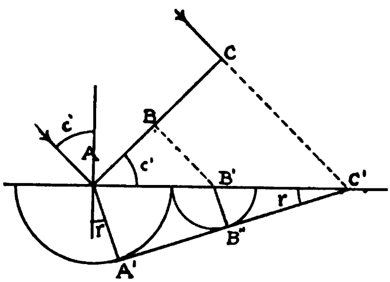


FIG. 255.

Consider a plane wave AC incident obliquely on the smooth plane surface of separation between air and another transparent medium (Fig. 255), the velocity in air being V_1 and that in the second medium V_2. A spherical wave will diverge from the point A into the second medium when the disturbance reaches that point, and later other spherical waves successively diverge from B' and C'. While the wave travels in the first medium a distance $CC' = V_1 t$ the wave from A will travel the distance $AA' = V_2 t$ in the second medium. The disturbance from B will in the same time travel a distance $BB' + B'B'' = (V_1 + V_2)t/2$, if B is half way between A and C. Since $B'B'' = \frac{1}{2}AA'$, a tangent plane can be drawn from C' to the two circles with centers at A and B'. It is easily shown by this method that the waves from all points in the original wave front will be tangent to the same plane, the new wave front in the second medium. Further,

$$CC' = AC' \sin i = V_1 t$$
$$AA' = AC' \sin r = V_2 t$$

Therefore,

$$\frac{\sin i}{\sin r} = \frac{V_1}{V_2} = n$$

The physical significance of the constancy of the sine ratio discovered by Snell thus becomes apparent. The student should always think of the index of refraction as being the ratio of the

velocities of light in the two media, rather than as the ratio of the sines of two angles. The latter mode of statement conveys no clear physical idea, and, moreover, seems to break down in the case of normal incidence.

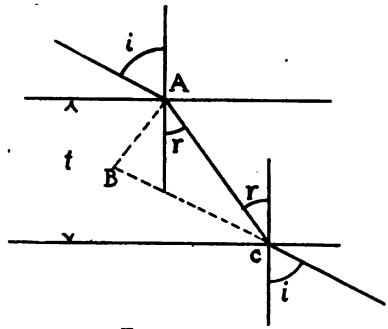


FIG. 256.

443. Medium with Parallel Surfaces. An incident pencil will be deflected in one direction on entering the second medium of thickness t and an equal amount on re-entering the first medium, as shown in Fig. 256. The course of the pencil will then be parallel to its original direction, but there will be a lateral displacement AB.

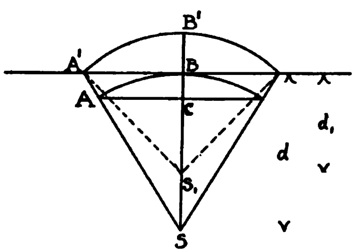


FIG. 257.

Image due to Refraction at Plane Surface. When an object is viewed normally to the boundary (Fig. 257) there is no lateral displacement, but only an apparent change in distance. Waves from an object S at a distance d below the surface of the medium travel with a velocity V_2 to the point B, where the vertex of the wave enters air, in which the velocity is V_1. The disturbance then travels a distance BB' in air while another portion of the wave still within the second medium travels the distance $AA' = V_2t$. The center of curvature of the emergent wave is at S_1, a distance d_1 below the surface. There is a virtual image of the source at this point. If the cone has only a small divergence, $AA' = BC$, the sagitta of the wave in the refracting medium, BB' is that of the wave in air, and $d = AS$ and $d_1 = AS_1$ their respective radii of curvature; hence, from the relation previously used (§ 432).

$$AA' = y^2/2d = V_2 t$$

Therefore,

$$BB' = y^2/2d_1 = V_1 t$$

$$\frac{d}{d_1} = \frac{V_1}{V_2} = n, \quad \text{or} \quad d = nd_1$$

The angle of the cone of light entering the eye is limited by the size of the pupil, and is, therefore, very small, so that the use of the above method is justified. The apparent depth of the object below the surface is $d_1 = d/n$. There is an apparent displacement toward the observer amounting to $(d - d_1) = (n - 1)d/n$. It is thus made clear why the depth of a pond appears to be less than it actually is, and why objects immersed in water appear to be shortened. Since the index of refraction is about 1.33, a pond six feet deep seems to be only about four and a half feet in depth.

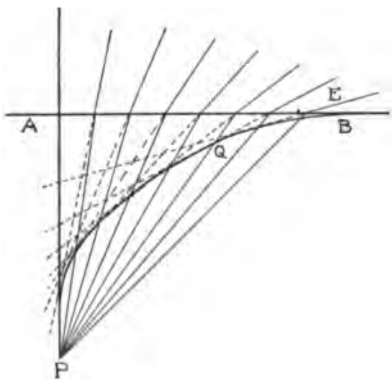


Fig. 258.

If the cone is wide there is considerable aberration, as shown in Fig. 258. This is not apparent to the unaided eye, which limits the aperture of the effective pencil, except through a slight lateral displacement (the image being at Q if the eye is at E), but may be made evident by using a large condensing lens of small focal length, which will take in a large divergent pencil, and convert the virtual image of the object into a real image equally affected by aberration.

444. Prism. If light waves pass through a transparent medium bounded by plane surfaces which are not parallel, the deviation of the incident pencil on entering the first surface is not exactly compensated on emerging from the second surface. If the source is at S (Fig. 259) the image, or center of curvature of the wave within the prism, is at S_1 and that of the emergent wave is at S_2. To

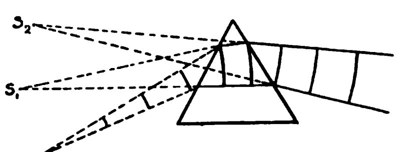


Fig. 259.

determine the deviation of the pencil and the positions of the foci S_1 and S_2 it is convenient to follow the course of given wave normal or "ray." The intersection of pairs of such rays will fix the position of the desired foci or centers of curvature of the waves.

445. Deviation—Minimum Deviation. The total deviation of a given ray is $D = D_1 + D_2$ (Fig. 260).

$D_1 = i_1 - r_1$; $D_2 = i_2 - r_2$;
$D = i_1 + i_2 - (r_1 + r_2)$

But $r_1 + r_2 = A$, since $B + A = 180° = B + r_1 + r_2$.

Therefore, $D = i_1 + i_2 - A$

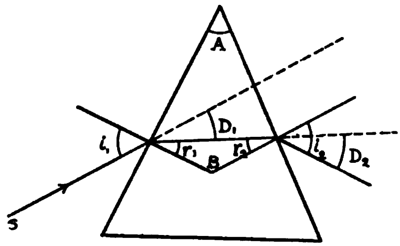


FIG. 260.

It is easily shown, experimentally or mathematically, that D has a minimum value when $i_1 = i_2$, in which case the incident and emergent ray are symmetrical with respect to the refracting angle of the prism. In this case

$$i_1 = i_2 = \frac{D + A}{2}$$

$$r_1 = r_2 = A/2$$

Therefore,

$$n = \frac{\sin i}{\sin r} = \frac{\sin \frac{D + A}{2}}{\sin \frac{A}{2}}$$

This relation is commonly used for determining the index of refraction of substances in the prismatic form. As the index is not the same for different colors, it is evident that the prism can set at the angle of minimum deviation for only one color at a time.

446. Focal Lines. If the source is a point S it is evident that the sections of the transmitted pencil in planes at right angles to the refracting edge will have their vertices in a line containing S_1 at a distance v_1 from the prism (Fig. 261); but a section parallel to the refracting edge will have its vertex in a line containing S_2 at right angles to the line at S, and at

a distance v_2 from the prism, v_2 being slightly less than u (Fig. 262). The displacement in the latter case is that due to a parallel-sided slab of thickness equal to that of the section of the prism considered. The resultant effect is that the emergent pencil appears to proceed from virtual focal lines F_1 and F_2 at distances v_1 and v_2 from the prism (Fig. 263). These

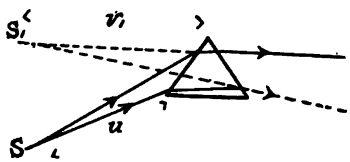


FIG. 261.

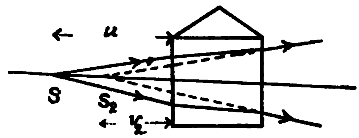


FIG. 262.

lines fuse together in a point image when the pencil passes through the prism at the angle of minimum deviation when $v_1 = v_2 = u$ nearly. If a converging lens be placed beyond the prism real focal lines will be formed.

447. Dispersion of Color. The index of refraction of any given substance varies with the color, or wave length; consequently the deviation caused by a prism will not be the same for all colors.

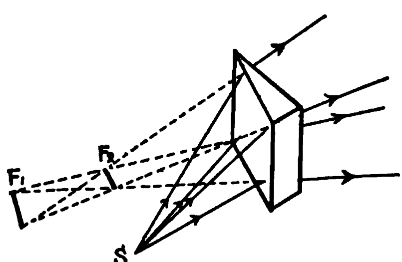


FIG. 263.

Consider a narrow source S such as an illuminated slit parallel to the edge of the prism (Fig. 264). If the source emits red light alone, a virtual red image of the slit is observed at R. If green and violet light are also emitted, a green and a violet image are seen at G and V. Real images of these colors may be formed at R', G', and V' by a lens. This separation of the colors is called *dispersion*. If the source emits waves of an infinite number of lengths included between the red and the violet, the infinite number

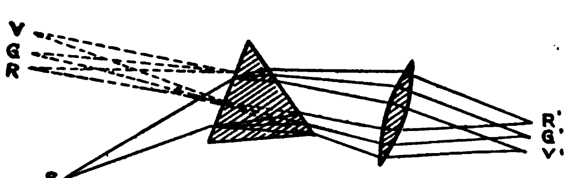


FIG. 264.

of partially overlapping images of the slit will form a continuous spectrum. If the slit is wide the different colors will greatly overlap, and the spectrum is said to be impure. There is less overlapping when the slit is narrowed; but, since no slit can be made infinitesimally narrow, it is manifestly impossible to obtain a perfectly pure spectrum.

448. Fraunhofer Lines. If a wide slit illuminated by sunlight is used, a continuous spectrum is observed, apparently like that given by a candle flame. Such a spectrum was observed by Newton. If, however, the slit is very narrow, it will be seen, as found by Wollaston in 1802, that a number of fine dark lines parallel to the slit cross the spectrum. He observed a virtual spectrum by looking directly through a prism at an illuminated slit. Fraunhofer, about 1815, by the use of better prisms, and by forming a real image of the spectrum with a lens, was able to find several hundred of these lines, which are now usually referred to as Fraunhofer lines. It is evident from this that the solar spectrum differs from that of a candle in not being absolutely continuous. The dark gaps in the position of different colors show the absence of corresponding images of the slit, and therefore the absence of these colors in the sunlight. In the section on Absorption (511) it will be shown that these dark lines are due to the absorption of light of definite wave lengths by vapors in the solar atmosphere. Luminous gases or vapors emit a limited number of colors, giving a discontinuous spectrum of isolated lines or slit images; and in many cases these vapors are opaque to the particular radiations which they emit when luminous.

The Fraunhofer lines may be used as reference points in measuring indices of refraction of prisms for different colors. The more prominent lines were labeled by Fraunhofer with letters of the Roman and Greek alphabets. Some of the more important of them are the *A* line (really a group of fine lines), due to absorption by the earth's atmosphere; the neighboring *D* lines, due to sodium vapor; the *F* line, due to hydrogen; the *H* and *K* lines, due to calcium. These lines are shown in the reproduction of the solar spectrum (Fig. 1, Plate I, p. 454).

449. Dispersive Power. The deviation of a particular color by a prism increases with the index of refraction. The angular

separation or dispersion between two colors depends on the difference between their respective indices of refraction. If a prism has a very small refracting angle, the angles of incidence, refraction, and emergence of a given pencil transmitted at the angle of minimum deviation will likewise be small, and the sines of these angles may be considered as equal to the arcs; consequently,

$$n = \frac{\sin\left(\frac{A+D}{2}\right)}{\sin\frac{A}{2}} = \frac{A+D}{A}$$

Therefore,

$$D = (n - 1)A$$

If D_1 and D_3 are the deviations of two given colors, the Fraunhofer lines C and F, for example, and D_2 that of an intermediate color halfway C to F, $D_3 - D_1$ is the angular dispersion of the extreme colors and D_2 is the mean deviation of the spectrum of angular width $D_3 - D_1$. The dispersive power d of the prism is the ratio of the angular dispersion of the two colors to their mean deviation, or

$$d = \frac{D_3 - D_1}{D_2} = \frac{(n_3 - 1)A - (n_1 - 1)A}{(n_2 - 1)A} = \frac{n_3 - n_1}{n_2 - 1}$$

Newton assumed that the ratio of dispersion to the deviation, or the dispersive power, is the same for all substances, but Dolland,

	n_D	$n_F - n_C$	$\frac{n_F - n_C}{n_D - 1}$
Water	1.3330	0.0060	0.0180
CS_2	1.6303	.0345	.0547
Ether	1.3566	.0052	.0149
Alcohol	1.3597	.0062	.0174
Crown glass	1.5160	.0073	.0141
Light flint glass	1.5718	.0113	.0197
Heavy flint glass	1.7545	.0274	.0363
Very heavy flint glass	1.9625	.0488	.0507
Quartz	1.5442	.0078	.0129
Diamond	1.4173	.0254	.0179
Iodide of silver	1.1816	.1256	.1063
Air (0°C., 760 mm.)	1.00024289	.00000295	.0121
H	1.00014294	.00000195	.0136
CO_2	1.00044922	.00000460	.0102

in 1757, showed that this is by no means the case. Two different prisms may have the same value of $n_2 - 1$, but very different values for $n_2 - n_1$, or conversely.

The table on the preceding page shows the dispersive power between the *C* and *F* lines for some substances, the mean deviation being that corresponding to the *D* lines. There are great differences between the refractive and dispersive powers of different specimens of glass.

450. Irrational Dispersion. If for any pair of colors n_2 and n_1 the ratio $(n_2 - n_1)/(n_2 - 1)$ were the same for all substances the spectra formed by different prisms would all be alike in the distribution of colors, and one spectrum could be simply a larger or smaller copy of any other. It is found, however, that this ratio is not constant, so that dispersion by a prism in general is said to be *irrational.* It is possible, for example, to make a prism of crown glass and one of flint glass which will give spectra of equal length between the lines *A* and *K*; but it will be found that the positions of the other Fraunhofer lines do not coincide in the two spectra, as they would if the dispersion were rational.

The following table showing the differences between the refractive indices of various substances for the *A, D, F,* and *G* Fraunhofer lines illustrates irrationality of dispersion. It will be seen that the ratio $(n_F - n_D)/(n_D - n_A)$, for example, is not the same for the different substances.

	$n_D - n_A$	$n_F - n_D$	$n_G - n_F$	$\dfrac{n_F - n_D}{n_D - n_A}$
Crown glass............	0.00485	0.00515	0.00407	1.062
Heavy flint glass......	.01097	.01271	.01062	1.158
Water..................	.00409	.00415	.00344	1.015
CS_2.....................	.01898	.02485	.02446	1.309

Although as a general rule the index of refraction increases as wave length diminishes, there are exceptions, as described under the head of anomalous dispersion (§ 552).

451. Achromatic and Direct Vision Prisms. The unequal dispersive power of different substances is utilized for making prismatic combinations for producing deviation with very little dis-

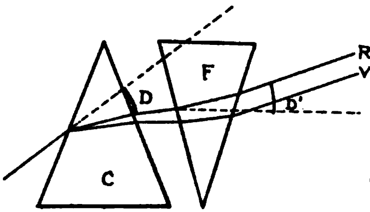


FIG. 265.

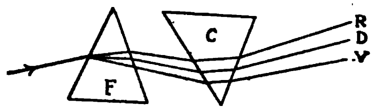


FIG. 266.

persion (Fig. 265), or dispersion without deviation of the spectrum as a whole (Fig. 266). These two types are respectively called *achromatic* and *direct-vision* prisms.

LENSES.

452. **Lenses** are transparent bodies, generally with spherical surfaces, which form images by changing the divergence of light waves. The ordinary types of single lenses are shown in Fig. 267. The first three forms, known as double-convex, plano-convex, and concavo-convex, are thicker at the center than at the edges. If surrounded by a less refractive medium, the central portion of the incident wave is more retarded than the edges by these lenses, and the curvature of the wave is diminished or reversed in direction. These lenses have, therefore, a convergent effect. They are

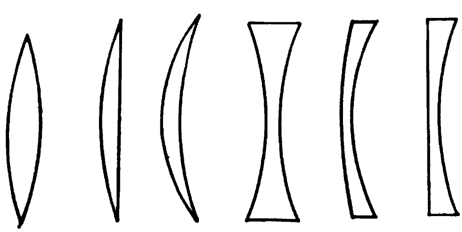


FIG. 267.

called *convex* or *converging* lenses. In the second group, embracing the double-concave, plano-concave, and convexo-concave lenses, the edges are thicker than the center, so that the outer portions of an incident wave are more retarded than the center. The curvature of the wave is increased and the lenses have a divergent effect. Such lenses are called *concave* or *divergent* lenses. If the two

types of lenses are placed in a more refractive medium there is a reversal of these effects.

453. Equivalent Air Path or Reduced Optical Path. At a given instant a wave front is in a given position; later it will be in a different position, and may have its orientation and curvature greatly modified by reflection or refraction. The one condition that must always be fulfilled, if the wave is to preserve its identity, is that the time required for the disturbance to travel from a point in the original wave front to the corresponding point in the new wave front is the same for all parts of the wave. For example, the disturbance traveling from S by the path $SPQS_1$ (Fig. 268) reaches S_1 at the same time as the disturbance leaving S at the same instant and traveling along $SAES_1$. The latter has been sufficiently retarded by passing through a greater thickness of glass to compensate for the greater distance $SPQS_1$. Similarly, the time required for the wave to travel from P to Q is the same as that from B to D. In comparing the distances traversed in equal times in different media, account must be taken of the velocity of light in the respective media. For example, in Fig. 268, $PQ = V_1t$; $BD = V_2t$. Therefore, $PQ = (V_1/V_2)BD = nBD$. If BD is the distance actually traversed in a medium of refractive index n, the *equivalent air path* or *reduced optical path* is nBD.

454. Conjugate Focal Relations. Consider the case of a double-

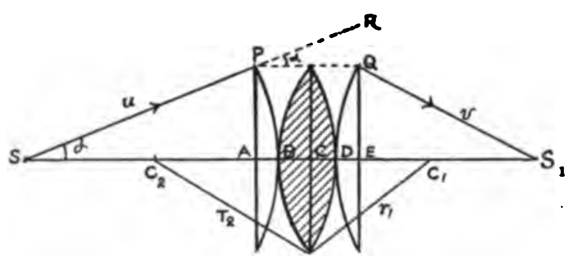


FIG. 268.

convex lens of refractive index n surrounded by air, the refractive index of which may be taken as unity. Let the radius of curvature of the first surface of the lens be r_1, that of the second r_2 (Fig. 268). Let u be the distance of the source from the lens. PB is a

section of the incident wave front of radius u, and QD that of the emergent wave front, of radius v.

The disturbance actually travels radially from P to R, but if a is very small, the path may be assumed to be PQ without appreciable error. Placing the optical path through the center of the lens equal to this distance, we have

$$PQ = AB + BC + CD + DE = n(BC + CD),$$

or

$$AB + DE = (n - 1)(BC + CD)$$

Substituting reciprocal radii of curvature for sagittæ (§ 433), this becomes

$$\frac{1}{u} + \frac{1}{v} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{1}{f}$$

In this u, v, r_1, and r_2 are considered as mere lengths, that is, numerical quantities without sign. As we shall later treat them as algebraical quantities with signs, it may be noted that, in the above case, u and v are both measured in the direction in which the light proceeds. It should also be noticed that the refraction at the first surface makes the wave less divergent, that is, it tends to converge it toward the opposite side. The same is true of the second surface. Hence both surfaces may be described as converging surfaces. If the curvature of either surface were opposite to its direction in the double-convex lens, it would be a diverging surface.

If the source of light is at an infinite distance, that is, if the incident waves are plane, $u = \infty$ and $v = f$. Hence f is the distance of the point called *the principal focus*, to which the lens converges plane waves.

In the case of a double concave lens, of thickness $t = CD$ along the axis (Fig. 269), if the incident wave front is PB and the emergent wave front QF, put the optical path BF equal to the optical path PQ (assumed parallel to the axis, since a is small),

$$AC - AB + n(CD) + DE + EF = n(AC + CD + DE)$$

$$\therefore AB - EF = (n - 1)(-AC - DE)$$

Substituting radii of curvature for sagittæ,

$$\frac{1}{u} - \frac{1}{v} = (n - 1)\left(-\frac{1}{r_1} - \frac{1}{r_2}\right) = -\frac{1}{f}$$

The above formula would be identical with that for the double convex lens if the signs of v, r_1, and r_2 were reversed. Now each of these quantities, in the case of the double concave lens, is actually measured in the opposite to the direction in which it is measured in the double convex lens. Hence, if we take the latter direction as positive for each, and so treat all of them (and v) as algebraical quantities, the two formulae will become identical.

From Fig. 269 it is evident that incident light is made more divergent by the lens. In fact both surfaces are diverging sur-

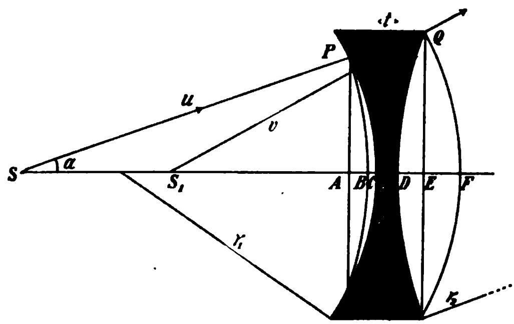


FIG. 269.

faces. The significance of the negative sign of f in the above formula is that the principal focus is virtual, its distance from the lens being measured in a direction opposite to that in which the light actually travels.

By applying the same method to the other types of spherical lenses it will be found that the general solution of all cases is the formula

$$\frac{1}{u} + \frac{1}{v} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{1}{f}$$

provided we adopt the following rules regarding signs:

(*a*) *Distances measured from the source to the lens and from the lens in the direction traveled by the light are considered positive.*

(*b*) *The radius of curvature of a converging surface is considered positive, that of a diverging surface negative.*

(*c*) *The sign of f is the same as that of $1/r_1 + 1/r_2$.*

Negative values of v and f indicate that the light diverges from a virtual focus after passing through the lens. These conventions are consistent with those of § 433.

Some readers may find the following alternative method of stating these rules more convenient.

(*a*) Consider u as positive when it is on the same side of the lens as the incident light and v as positive when it is on the opposite side.

(*b*) Consider the radius of curvature of a convex surface of a lens to be positive, that of a concave surface negative.

(*c*) Consider f to be positive for a converging, negative for a diverging lens.

The following cases arise when f is positive:

When $u = \infty$, $v = f$, the principal focal distance.

When $u > f$, v is positive and there is a real conjugate focus.

When $u = f$, $v = \infty$. The transmitted beam is parallel.

When $u < f$, v is negative and greater than u for all positive values of u, and there is a virtual conjugate focus.

When f is negative,

$$v = -\frac{uf}{u+f}$$

The student may as an exercise determine what modification of the formula showing conjugate focal relations is made necessary if the lens is surrounded by a medium with index of refraction n_1 or has a medium with index n_1 on one side and one with index n_2 on the other. He may also study the focal properties of a single spherical surface bounding two media of indices n_1 and n_2.

455. Axes of Lens. The line passing through the centers of curvature of the surfaces of a lens is called the *principal axis*. In every lens there is a point on the principal axis, called the *optical center*, which has the property that no ray passing through it is deviated in direction, although there is more or less displacement, depending on the thickness of the lens.

The existence of this point may be shown thus: Let two parallel radii of curvature r_1 and r_2 (Fig. 270) be drawn to the two surfaces of a lens. Since

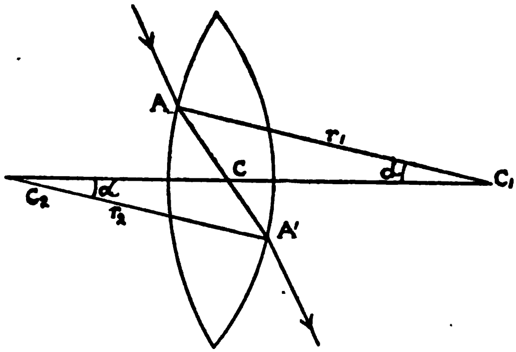


FIG. 270.

the two plane elements of the lens A and A' are parallel, being respectively perpendicular to two parallel lines, the refracted ray AA' is propagated in a medium with parallel sides and emerges parallel to its original direction. Since the triangles ACC_1 and $A'CC_2$ are similar,

$$\frac{r_1}{CC_1} = \frac{r_2}{CC_2}$$

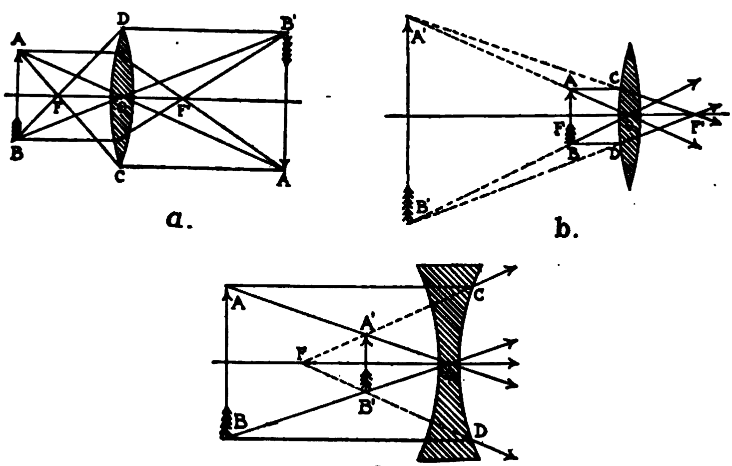


FIG. 271.

This is true whatever may be the value of the angle α, therefore C is a fixed point, the optical center of the lens. All ray paths which pass through this point are called *secondary axes.* In the case of a thin lens, the center of the lens and the optical center may usually be regarded as coincident.

456. Images by Lens. The image of A (Fig. 271, *a, b, c*) must lie on the secondary axis AA', that of B on the secondary axis BB'. Rays drawn parallel to the principal axis from the points A and B pass through the principal focus F and intersect the lines AA' and BB' at the points A' and B', which determine the position and magnitude of the image. Since the point A' lies above the principal axis when the image is on the same side of the lens as the object, and below it when the image is on the other side of the lens, it is evident that all virtual images formed by a single lens are erect, all real images inverted.

457. Magnification. Since AB and $A'B'$ subtend equal angles from the center of the lens (the angle included between the secondary axes AA' and BB') it is evident that their relative sizes are proportional to their respective distances from the lens, or

$$\frac{AB}{A'B'} = \frac{u}{v}$$

458. Spherical Aberration. In deriving the formula for the conjugate focal relations of lenses it has been tacitly assumed that

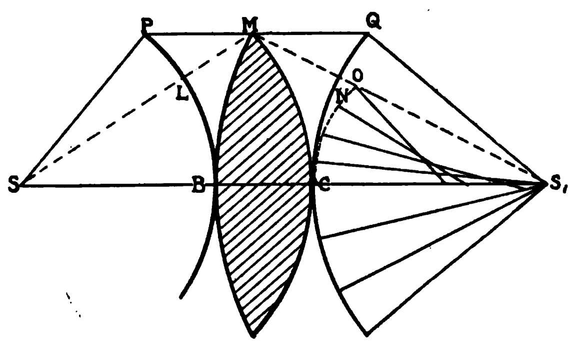


Fig. 272.

the emergent wave is spherical. With lenses of small aperture this is shown by experience to be practically true; but when the

aperture becomes large there is noticeable spherical aberration. This is illustrated by Fig. 272.

While the central part of the wave travels from B to C the edge of the wave will travel along LMN the distance $LMNO = PQ = nBC > LMN$. It is evident, therefore, that the edge of the emergent wave (represented by the dotted curve) will pass through O instead of N, and have a greater curvature toward the axis than if the wave were spherical, with S_1 as a center. The rays, instead of converging to S_1, as shown in the lower half of Fig. 272, will cross each other as shown in the upper half, being enveloped by a caustic surface instead of by a right cone.

459. Correction of Spherical Aberration. If the rays passing through the edge of a lens are stopped by a diaphragm which permits only the central portion of the incident pencil to pass the spherical aberration will be greatly reduced. It is also possible to grind surfaces slightly differing from a spherical form, so that for a given pair of conjugate focal distances the emergent wave is truly spherical. Such lenses are called *aplanatic*. In some cases when the conjugate focal distances differ greatly, spherical aberration may be reduced by making the two surfaces of the lens of different curvatures. Consider, for example, a

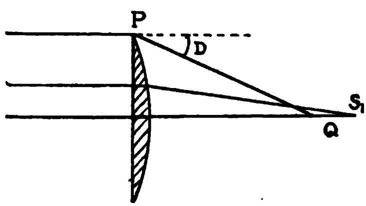


FIG. 273.

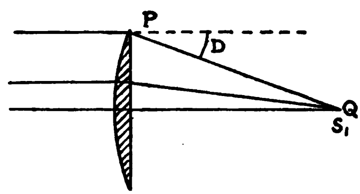


FIG. 274.

plano-convex lens of great aperture (Figs. 273, 274) first with the plane then with the convex face toward a source so distant that the incident light is parallel or nearly so. If we consider the deviation of the ray PQ in each case, it is evident, on recalling the condition for minimum deviation by a prism, that in the second case the angle D will be less than in the first, because the refracting edge of the lens is then more nearly in the position with respect to the incident and emergent rays which gives minimum deviation, and consequently the nearest approach of the ray PQ to the focus S_1.

One form of thick lens of great angular aperture commonly used as part of microscopic objectives is almost entirely free from spherical aberration. Suppose that a pencil of light converging to S falls on a transparent sphere of radius R and is refracted to the focus S_1 (Fig. 275). Using a well-known geometrical principle, we may write

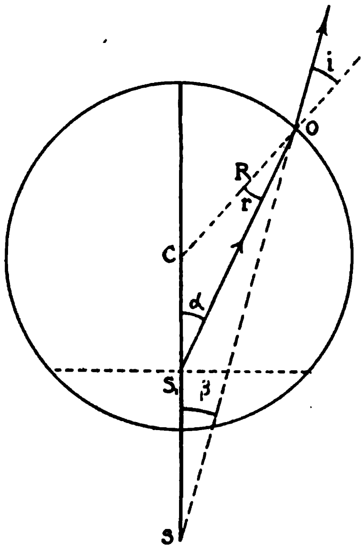


FIG. 275.

$$\frac{R}{\sin \alpha}=\frac{CS_1}{\sin r}; \quad \frac{R}{\sin \beta}=\frac{CS}{\sin i}$$

Therefore,

$$CS_1=R\frac{\sin r}{\sin \alpha}, \quad CS=R\frac{\sin i}{\sin \beta}$$

The angle at C is common to the two triangles CSO and CS_1O, hence $r+\alpha=i+\beta$, and if $\beta=r$, $\alpha=i$. In this case, if n be the index of refraction of the sphere,

$$CS_1=R/n; \qquad CS=nR$$

This is an exact relation, no matter how large the angle i may be, so that an object in the lens at S_1 would have a virtual image at S entirely free from aberration. The same is practically true if the segment of the sphere below S_1 is removed and the object placed in contact with the surface. In practice this lens is often a hemisphere, the object being placed at such a distance below the plane side that its virtual image formed by refraction at the plane surface corresponds in position to the point S_1. There is some aberration in this case due to refraction at the lower surface.

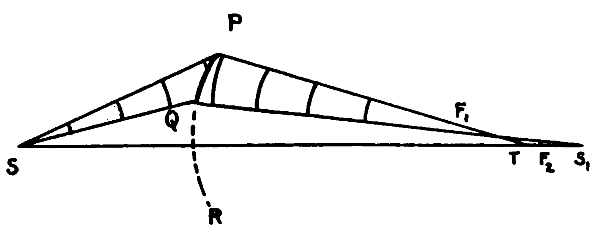


FIG. 276.

460. Focal Lines. If a pencil of light falls obliquely on a converging lens, instead of a point image two real focal lines will be formed, like those due to a concave mirror. If the lens is divergent, these focal lines

will be virtual. The formation of these lines by a converging surface is made clear by considering the effect of a single refracting surface PQ, imagining it to be extended to R, so that SS_1 is the principal axis (Fig. 276). The rays transmitted through the actual refracting surface PQ, will, by reason of spherical aberration, pass through a narrow arc through F_1 with its center on SS_1. This is the primary focal line. These rays will all intersect the axis SS_1 between S_1 and T. The normal cross-section of this pencil is a narrow lemniscate-shaped region at F_2, the secondary focal line, at right angles to the primary focal line. The second refracting surface of the lens will modify but not change the general character of this result.

461. Cylindrical Lens. The effect of such a lens is like that of a cylindrical mirror. A point source S has two linear images, as shown in Fig. 277, one AB parallel and the other $A'B'$ at right angles to the axis of the lens. The image AB parallel to the axis is at a distance given by the relation

$$\frac{1}{u}+\frac{1}{v}=(n-1)\left(\frac{1}{r_1}+\frac{1}{r_2}\right)$$

and may be either real or virtual; the other, $A'B'$ is virtual and is as far behind the lens as the image given by a parallel-sided plate of the same thickness as the lens. Any lens with different curvatures in planes at right angles to each other will give similar focal lines or astigmatic images.

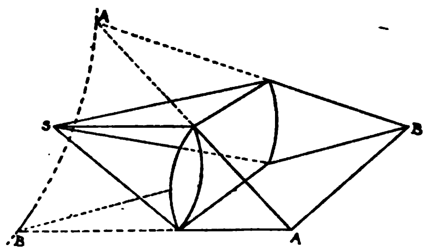


FIG. 277.

462. Combination of Lenses. If the lenses are thin, with principal focal lengths f_1 and f_2, and so close that distance between them may be neglected,

$$\frac{1}{u}+\frac{1}{w}=\frac{1}{f_1};\quad -\frac{1}{w}+\frac{1}{v}=\frac{1}{f_2}$$

If w, the focal distance conjugate to u, is positive with respect to the first lens, it is negative with respect to the second. Therefore,

$$\frac{1}{u}+\frac{1}{v}=\frac{1}{f_1}+\frac{1}{f_2}=\frac{1}{f}$$

This expression is generally true for either converging or diverging lenses if the proper signs are given to f_1 and f_2.

463. Chromatic Aberration. Since

$$\frac{1}{f} = (n - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = (n - 1)K$$

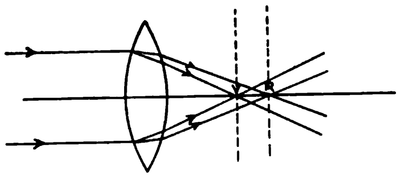


FIG. 278.

it is evident that the principal focal distances are different for different colors, being less for violet than for red (Fig. 278). There is no way to remedy this defect in a single lens, but it may be greatly reduced by a suitable combination of lenses.

464. Achromatic Combinations. By combining two or more lenses of different dispersive powers, two or more given colors may be brought to the same focus, just as prisms may be com-

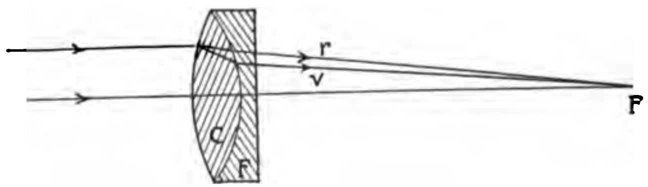


FIG. 279.

bined to give deviation without dispersion (Fig. 279). If two lenses are used, for each color

$$\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$$

if the lenses are in contact. If we wish to combine the two colors corresponding to the C and F lines, f must be the same for both.

If n_F' and n_C' be the refractive indices of the first lens, n_F'' and n_C'' those of the second,

$$\frac{1}{f} = \frac{1}{f_F'} + \frac{1}{f_F''} = (n_F' - 1)K_1 + (n_F'' - 1)K_2$$

$$= \frac{1}{f_C'} + \frac{1}{f_C''} = (n_C' - 1)K_1 + (n_C'' - 1)K_2$$

therefore,

$$(n_F' - n_C')K_1 = (n_C'' - n_F'')K_2$$

The values of $K_1 = 1/r_1' + 1/r_2'$ and $K_2 = 1/r_1'' + 1/r_2''$ may be arbitrarily chosen to satisfy this relation. Since $n_F > n_C$ it is evident that K_1 and K_2 must be of opposite sign, so that either f_1 or f_2 must be negative. If f_2 is negative and greater than f_1, f is positive and the lens is convergent. If f_2 is negative and less than f_1 the combination is divergent. Usually the positive lens is of crown glass, the negative of flint, and they are shaped to fit close together, so that $r_2' = r_1''$ and often $r_2'' = \infty$ (Fig. 279). To consider a practical example, suppose we wish to construct a convergent lens of crown and heavy flint glass which will combine the F and C lines, with a focal length of 20 cm., the flint glass lens being plano-concave. Substituting numerical values for the indices, as given in § 449,

$$-\frac{K_1}{K_2} = \frac{n_F'' - n_C''}{n_F' - n_C'} = \frac{.0274}{.0073} = 3.75$$

But $K_1 = 1/r_1' + 1/r_2'$, $K_2 = -1/r_2'$, therefore,

$$-\frac{K_1}{K_2} = \frac{r_1 + r_2}{r_1} = 3.75\,; \quad r_2 = 2.75 r_1$$

The values of r_1 and r_2 corresponding to any given value of f can be calculated. By using three lenses of different dispersive powers three colors may be combined, but for most purposes a combination of two lenses is sufficient.

Chromatic aberration may also be reduced by using two lenses of the same index of refraction at a certain distance d from each other. To take a specific case, if the second lens is placed at a distance from the first equal to its own focal length, the rays of different colors which diverge from each other at the first lens will be made approximately parallel by the second. If an object is placed at the principal focal point F (Fig. 280) of the combination, a virtual image at infinity will be formed, and, as shown by the figure, the violet and the red images will subtend approximately equal angles α at the eye, and will, therefore, be superimposed on the retina.

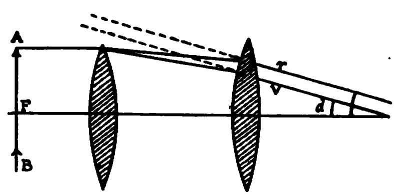


FIG. 280.

REFRACTION PHENOMENA.

465. Total Reflection. If a ray of light travels from a more to a less refractive medium, the angle of emergence i is greater than the angle of incidence (which, being in the more refractive medium, may still be called r for consistency). Since $\sin r = (\sin i)/n$, and

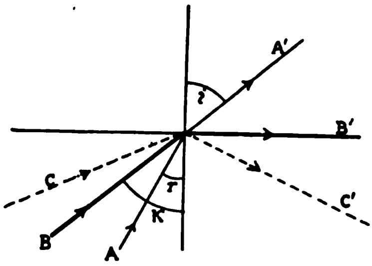


Fig. 281.

since i has a maximum limit of 90 degrees, r has a maximum limit k such that $\sin k = 1/n$. No pencil incident on the boundary at a greater angle than k, the *critical angle*, can emerge. It will, therefore, be totally reflected (CC', Fig. 281). Since $\sin k$ varies inversely as the index of refraction, the critical angle is different for different colors. Violet will first be subject to total reflection as r increases, and finally the red.

A parallel-sided plate cannot be used to show total reflection, since any pencil entering such a plate must emerge at the same angle. Bodies of prismatic form are best adapted for the purpose. The effects may be seen by looking down at the side of a glass containing water or a test-tube sunk in water. A fish can see

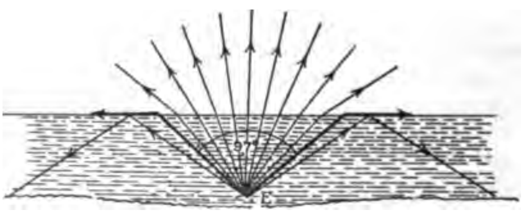


Fig. 282.

objects throughout the space above the water, but he sees them through the limited cone of angle $2k = 97°$ (Fig. 282), arranged around a circle, with tops pointing inward.

Some values of k are given below:

Water	48° 36′	Quartz	40° 22′
Crown glass	43° 2′	Diamond	24° 26′
Flint glass	37° 34′		

The smaller the critical angle of a jewel with regular facets, the greater the proportion of light totally reflected by it. This explains the great brilliancy of the diamond.

466. Transition Layer. It seems quite possible that the change of index of refraction at the boundary is not abrupt, but that there is a transition layer *t* due to interpenetration of the two media, or occlusion at the surface, causing a gradual change in the index. If this be the case, total reflection may be considered as altogether due to refraction. When the angle of incidence is equal to or greater than *k* the wave front in the transition layer will swing around and become normal to the surface (Fig. 283); then the lower edge will gain on the upper and the wave will swing

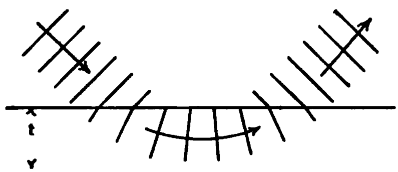


FIG. 283.

back into the first medium. If we consider an air film between two refracting media the two transition layers may encroach on each other (Fig. 284), in which case the lower edge of the wave will be retarded, and a part of it will pass into the third medium. It might be expected, therefore, that if the air film from which total reflection takes place is very thin total reflection will cease. This has been found to be the case. In the arrangement of prisms shown in Fig. 289 total reflection takes place from the hypothenuse of the first prism when the angle of incidence *i* on *AB* is sufficiently small, but some light will always be transmitted through the region surrounding the point of contact even where the air film has a measurable thickness. It is found that the thickness of the air film through

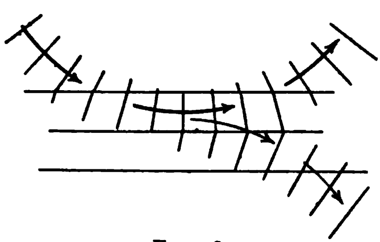


FIG. 284.

which transmission can occur (which may be considered as approximately the thickness of the transition layer) differs with the wave length and with the angle of incidence, and may reach several thousandths of a millimeter.

467. Mirage. Examples of the type of total reflection referred to above are found in the case of refraction by gases of varying density. This phenomenon is called mirage. The air above a furnace or a heated surface

such as a pavement exposed to the sun's rays rapidly increases in density and refractive power in going upward. If the line of vision forms a small angle with the surface distant objects are seen apparently reflected from the surface. The formation of one type of mirage is shown in Fig. 285. The object *AB* is viewed directly through the pencils *OA*, *OB*, while an in-

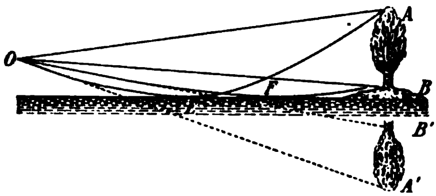


FIG. 285.

verted image *A'B'* is also seen, due to the refraction of the pencils *OEA*, *OFB*, by the heated air near the ground. This is one of several types of atmospheric mirage. Other types showing distortion or displacement of objects are due to local differences of temperature in the atmosphere, causing changes in density and refractive power. They are very easily seen by viewing objects at a grazing angle across heated surfaces. Similar effects are to be seen by looking through sheets of glass with irregular surfaces, or non-homogeneous mixtures of liquids, such as water with an excess of

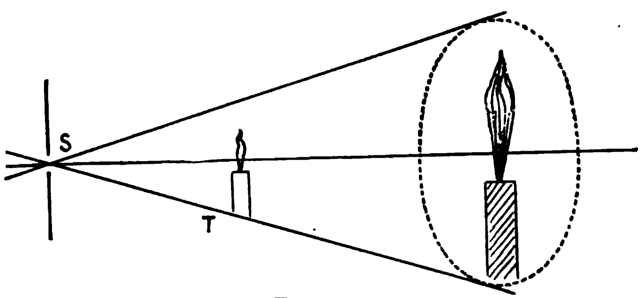


FIG. 286.

salt crystals at the bottom of the vessel, or with alcohol above and imperfectly mixed with it.

When the sun is near the horizon, the rays reaching the eye traverse strata of air of gradually increasing density, which cause them to bend downward. For this reason the sun is visible when it is actually below the horizon a distance about equal to its own diameter.

The scintillation of stars is due to a similar cause. Their apparent

direction and intensity are subject to rapid fluctuations as masses of air of varying density drift across the line of sight.

468. "Schlieren" Method. Many phenomena which are too feeble to be observed directly by the eye may be made visible by projection if they occur in the cone diverging from a point source of light. If sunlight be focused on the small hole S (Fig. 286) beyond which it diverges to a screen, a jet of coal gas, hydrogen, or carbon dioxide from the tube T will cast a clear image on the screen. The difference between the refractive power of the jet and that of the air will cause an alteration in the distribution of light on the screen, which will make the projected areas lighter or darker than the surrounding space. Ether vapor poured from a beaker, or ether vortex rings, may thus be made visible. If a bunsen burner be placed in the cone of light a beautiful representation of the flame and the currents of heated gases will be formed on the screen. This is a simple form of what is known as the "schlieren" method, due to Toepler, which was at first used to discover defects in lenses, but which has also been used by Toepler and by Wood to make visible projections or photo-

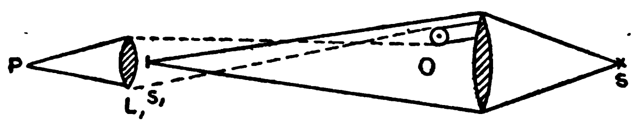


FIG. 287.

graphs of sound waves. Suppose an image of the source S (Fig. 287) to be formed at S_1. By means of a small screen at S_1 all or nearly all of the light may be cut off from the lens L_1. If in a small region O the index of refraction is slightly different from that of the air (or if O is a defective spot in the lens L) the light passing through this region will be made slightly more or less convergent than that of the rest of the pencil, and, as shown in the figure, will pass around S_1 and through L_1 and be brought to a focus at P, where the illumination will be increased or diminished in a region having the same form as O. If O be a cylindrical sound wave produced by an electric spark parallel to SS_1 it may be seen or photographically recorded by the flash of another spark at S.

469. The Rainbow is a bright arc showing the spectral colors, due to the sunlight reflected from raindrops. Sometimes several bows are seen, the inner or primary bow being always the brightest, and all being arcs of circles with centers on the prolongation of the line passing from the sun through the observer. The primary bow is violet on the inside, red on the outside; in the secondary bow the order of colors is reversed.

De Dominis showed that the rainbow is due to light refracted and internally reflected by raindrops; Descartes determined by a geometrical method the angles subtended by the various bows at the observer's eye, and Newton explained the color effects as being due to dispersion within the drop. Young proved that interference effects modify the distribution of the colors, and cause the supernumerary bows sometimes seen. Airy (1838) and Pertner (1897) developed an exact theory, taking account of the interference between rays which have experienced relative retardations within the drop, and showed that not only the distribution of colors, but also the angular radius and breadth of the bow are largely determined by

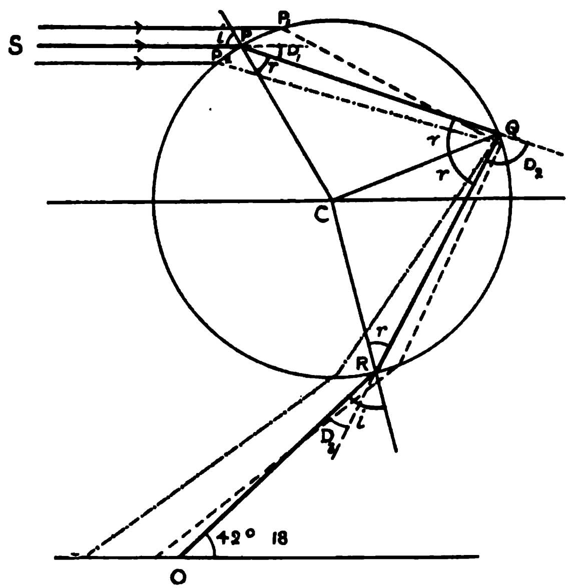


FIG. 288.

interference, these effects varying with the size of the drops. The approximate angular magnitude of the various bows may, however, be determined by the simpler geometrical method of Descartes, with little error in the case of large drops.

If parallel rays are incident on the upper half of a refracting sphere they will be in part refracted, internally reflected, and transmitted downward as shown in Fig. 288. Rays will also enter the lower half, and there will be multiple reflection within the sphere, but for the present we shall fix our attention upon the rays which reach the eye at *O* after one internal

reflection. As indicated by the course of the rays incident at P, P_1, and P_2, there is an angle of minimum deviation, below which no rays once internally reflected pass. All the rays emerging at nearly this angle are parallel or nearly so, and therefore their intensity varies little with distance from the drop, while rays emerging in other directions are widely divergent.

As n varies with the color, the minimum deviations are different for the various colors. In the primary bow the minimum deviation of the red is 137° 42′; of the violet, 139° 37′. In the secondary bow the corresponding angles are 230° 34′ and 233° 56′.

From Fig. 288 it appears that light will be received by the observer at O from all the raindrops lying in an arc subtending an angle 180°—D with the axis passing from the sun through the observer's eye. In the primary bow this angle is 42° 18′ for the red and 40° 23′ for the violet, so that the bow will be bordered with violet on the lower side, red on the upper. The secondary bow is due to rays incident on the lower half of the drop, twice internally reflected, and then transmitted downward, thus inverting the order of the colors. The angle subtended by this bow is D—180°, or 50° 34′ for the red, 53° 56′ for the violet.

An artificial rainbow may be made by causing a beam of sunlight to fall on a spherical vessel filled with water, through an opening in a screen. The interior of the circle reflected on the screen is illuminated by the scattered light which has been once reflected, while the space between the primary and secondary bow is quite dark.

INTERFERENCE.

470. Colors of Thin Plates. As previously noted, thin films of transparent substances, such as soap bubbles, layers of oil on water, and thin sheets of mica, show brilliant color effects in reflected white light, and complementary effects in the transmitted light.

Newton's rings are due to interference between the light waves reflected from the upper and lower surfaces respectively of a thin film of air of varying thickness enclosed between a glass plate and a convex lens of great radius of curvature. When white light is incident normally on such a combination the rings or "fringes" are circular and concentric with the point of contact. If the light is incident obliquely, the fringes are elliptical. These bands are violet on the inside and red on the outside; but at a short distance from the center they begin to overlap and conse-

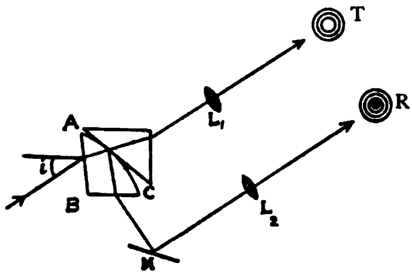


FIG. 289.

quently to grow indistinct. If monochromatic light is used a much larger number of bands may be seen. A good arrangement for projecting these rings and showing their complementary character is represented in Fig. 289.

A right-angled prism with a slightly convex hypothenuse face is pressed against the plane hypothenuse surface of another prism. A parallel beam of sunlight is incident on the side AB at an angle i sufficiently great to avoid total internal reflection. The reflected light passes out through the side BC and is reflected by the mirror M to a position R near or superimposed on the transmitted light T. The colors on the screen are visible without the use of a lens; but if lenses L_1 and L_2 are placed so that the film and the screen are at their respective conjugate focal planes the fringes will be magnified and at the same time made more distinct. It will be shown later (Fig. 290) that the interfering pairs of rays from a film of varying thickness intersect at or near the film and diverge, so that they must be again superimposed by a lens.

If the prisms are pressed closer together the rings will expand, because the loci of points of equal thickness have moved outward. If they are so close that the thickness of the air film at the point of so-called contact is small compared with the wave length of light, the central spot is dark in the reflected fringes, bright in the transmitted system. If the lenses form rings of equal size, T and R on being superimposed will give a uniform white background, showing the complementary character of the fringes.

It has been found impossible to produce interference effects between two pencils from separate sources, or from different points of the same source. There is no permanent concordance in phase relations, amplitude, or direction of vibration. As aptly expressed by Edser, "no interference phenomena could be produced by using two separate candle flames, any more than two brass bands playing different tunes in the same street could produce silence." We need, therefore, consider only one point of the source at a time. The effects of adjacent points will be simply superimposed on those of the first, without mutual interference.

Let LM and NO (Fig. 290) represent two opposite elements of surface of the air film producing Newton's rings, slightly inclined to each other and so small that they may be considered plane. A narrow pencil from the point S of an extended source is incident at A, where a small part is

reflected, the remainder being transmitted to B, where it is again subject to reflection and refraction. This process is repeated at C, D, E, etc., but the components become very weak after a few reflections. Owing to the inclination between the surfaces, the reflected pencils are not parallel, but intersect in the neighborhood of P_1, while the transmitted components appear to diverge from P_2. If the film increased in thickness toward the right, these points of intersection would be respectively on the opposite side of the film. The reflected and transmitted pencils may respectively be brought together again by the lenses L_1 and L_2, at the points S_1 and S_2. The eye or observing telescope must, therefore, be focused on P_1 or P_2 to get the most distinct effect. If the film is very thin these points practically lie at its surface. If either the thickness of the film or the angle between its surfaces is large the pencils are so distant from each other or so divergent that they cannot all enter the pupil of the eye. Under such conditions the bands become indistinct or vanish.

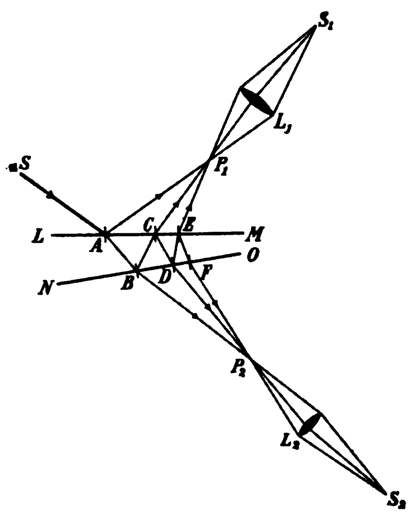


FIG. 290.

To get a general expression for the difference of path, let n be the index of refraction of the film, n_1 that of the surrounding medium, and for simplicity imagine the two surfaces to be parallel. To reach the wave front CP (Fig. 291) the light reflected from A travels the distance AP in the first medium while the interfering component has to travel the distance $AB + BC$ in the film. The wave would travel a distance $n_1\,AP$ in air while traveling the distance AP in the medium of index of refraction n_1 and the distance $n(AB + BC)$ in air while traveling the distance $AB + BC$ in the film (§ 453). Hence the equivalent difference of path in air, or the optical difference of path is

$$d = n(AB + BC) - n_1 AP.$$

But $AC \sin i = AP$; $AC \sin r = CQ$; therefore,

$$AP = CQ\frac{\sin i}{\sin r} = \frac{n}{n_1}\,CQ$$

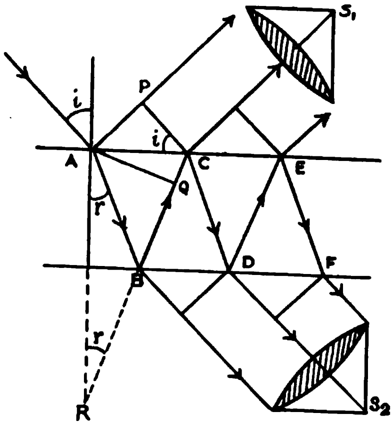


FIG. 291.

hence

$$d = n(AB + BC - CQ) = n(AB + BQ) = nAR \cos r = 2nt \cos r.$$

If the film is of air, $n = 1$, $d = 2t \cos i$ (if i is the angle of refraction in air). The effects at S_1 and S_2 are due to the superposition of all the components arising from multiple reflections within the films. Between successive pairs there is the same phase difference.

471. Phase Changes in Reflection. (See § 355.) Since $d = 2t \cos i$, there should be, so far as geometrical differences of path are concerned, reënforcement from the components reflected from the region of contact, where t is so small compared with the wave length of light that it may be ignored. As a matter of fact, the center of the reflected system of fringes is black. Young inferred by analogy that at the boundaries of different media light waves are subject to changes of phase similar to those observed in the case of material waves (§ 366) so that waves incident from air on a more refracting medium may behave like waves of sound reflected from a medium denser than air, while a light wave traveling in the opposite direction will behave like sound waves emerging from the free end of an organ pipe. The waves reflected from the upper surface of the air film pass from a more to a less refractive medium; at the lower surface the contrary is the case. If t is small compared with the wave length, there should be a difference of half a period introduced in the act of reflection, which will cause destructive interference. The transmitted components have a difference of phase of an entire period caused by two internal reflections, and therefore will be concordant. This would explain the black spot seen in the center of the reflected system of Newton's rings. It is also observed that soap films as they get thinner run through a brilliant series of colors when viewed by reflected light, finally becoming black just before they break.

In the arrangement described in § 470, if one prism is of crown glass, the other of flint, and if the interspace is filled with a liquid of intermediate index of refraction, such as oil of cloves, the central spot of the reflected system will be bright, that of the transmitted system dark. This confirms Young's theory.

When Newton's rings are produced by an air film, the condition for a

maximum of given wave length λ in the reflected light is (remembering that a loss of half a period in reflection is equivalent to a path difference of $\lambda/2$)

$$n\lambda + \tfrac{1}{2}\lambda = \frac{2n+1}{2}\lambda = 2t\cos i$$

and for a minimum, $n\lambda = 2t \cos i$.

In the transmitted light the maxima are given by

$$n\lambda = 2t\cos i$$

and the minima by

$$\frac{2n+1}{2}\lambda = 2t\cos i$$

In the above expressions n is the ordinal number of the rings counted from the center.

Film with Parallel Sides.—If the surfaces of a thin plate are perfectly plane and parallel the interfering rays are parallel, as shown in Fig. 291, and the eye or observing telescope must be focused for infinity to see the bands clearly. Since the difference of path between the components, $2nt \cos i$, varies with the angle of incidence, the phase relations will be different for rays reflected from different parts of the film, but will be the same for all rays reflected from the film at the same angle. The light from every point S_1, S_2, etc., of an extended source will be brought to separate points, S_1', S_2', etc., on the retina, so that there will be no overlapping of effects. The bands will in general be curved, their loci being given by $\cos i =$ constant. Such bands are sometimes known as Haidinger's fringes, or fringes of equal inclination.

472. Interference by Thick Plates. It is usually impossible to get interference effects by the use of a single wedge-shaped plate unless the inclination of the surfaces is very slight, because the interfering pencils will otherwise be too divergent to simultaneously enter the eye. If the surfaces of the plate are perfectly plane and parallel it is easy to obtain interference effects with monochromatic light with great differences of path between the components. The limit to the possible differences of path which may exist seems to be due to the lack of perfect homogeneity in the light from available sources, or to the probable fact that radiating centers emit detached wave groups corresponding to successive stimuli, these groups having different relations of phase, amplitude, and direction of vibration, so that waves of one group cannot interfere with those of another. Consequently the maximum

difference of path which can exist cannot exceed the length of such a train of waves.

473. Stationary Light Waves. If plane sound waves are reflected from a wall stationary vibrations are produced by interference between the incident and the reflected waves, resulting in the formation of nodal planes (§ 382) parallel to the wall and distant $\lambda/2$ from each other. Similar effects may be expected if plane waves of light are reflected normally from a mirror, but as the distance between the planes is only $\lambda/2$ and light waves are very short, it is difficult to verify their existence. Wiener (1889) did so by a very ingenious device. A glass plate *AB* (Fig. 292) was covered with a very thin photographically sensitive collodion film, and placed film downward over a silvered mirror *MN* with a very slight inclination between the two surfaces. After exposing the plate to a beam of light incident normally on the mirror it was developed, and dark bands were found in the film, running parallel to the line of intersection of the two surfaces, as indicated by the shading below *MN*. From the figure it is clear that the sensitive surface crossed the nodal and anti-nodal planes in such a manner as to produce this effect. At the points in contact with the mirror no effect was produced in the film. This proved the existence of a nodal plane at that surface, as in the case of the analogous sound experiment.

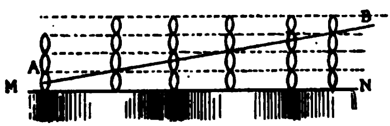


FIG. 292.

474. Lippmann's Color Photography. This method is dependent on the formation of nodal reflecting planes in a film by Wiener's process. A transparent sensitive film several wave-lengths thick is floated on mercury, and a colored image, say a spectrum, is thrown normally upon it. Stationary waves will be set up in the film, the distance between the nodal planes varying with the color (Fig. 293). After development the anti-nodal planes, which have been subject to the greatest light disturbance, will become partially opaque reflecting surfaces, and if white light falls on them the conditions for reënforcement of certain colors by reflection, as in the case of Newton's rings, will exist. Red will be most copiously reflected from the places where red light originally fell, and so with the other colors, so that the spectrum, diluted more or less with white light, will become visible by normal reflection of white light. If viewed at any other angle the colors will change, or become indistinct.

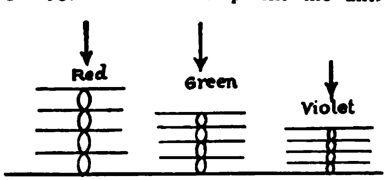


FIG. 293.

DIFFRACTION.

475. If light from a small source or aperture passes by the edge of an obstacle and falls on a screen it is found that the illumination gradually fades away in the geometrical shadow, while out-

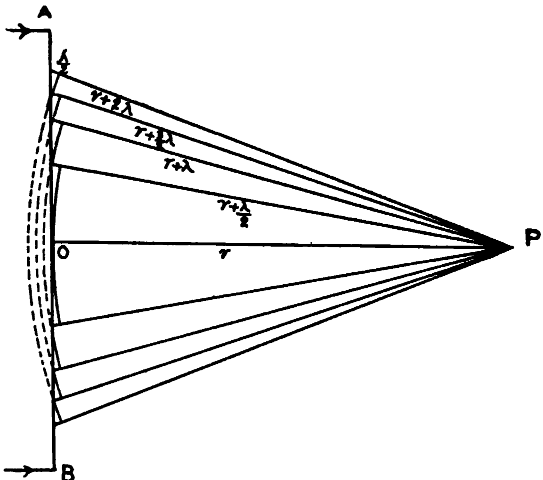


FIG. 294.

side the shadow a series of colored bands appears. If a card or knife blade is held between the eye and a distant source of light it will be found that the red light is most deflected into the shadow, the violet the least, so that a short spectrum is formed. Such phenomena are examples of what is known as **Diffraction.** Grimaldi, Newton, and Young observed these effects, but Fresnel was the first to explain them as the necessary consequence of the undulatory nature of light. They are, in fact, interference phenomena between wavelets coming from adjacent points of the same wave front.

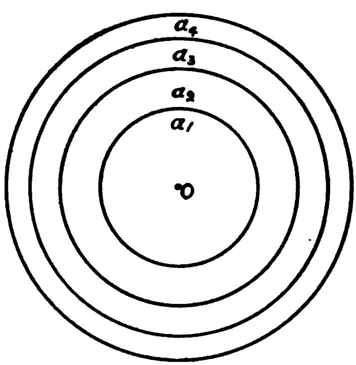


FIG. 295.

Let us find the effect of an extended wave plane front AB (Fig. 294) at the point P. In accordance with Huyghens' principle, the resultant effect at P may be regarded as the sum of the effects separately due to all the points in the wave front, each originating its independent set of wavelets. Waves of different length must be separately considered in this analysis. If $OP = r$, describe about P as a center spheres of radii $r + \lambda/2$, $r + \lambda$, $r + 3\lambda/2$, etc. These spheres will intersect the wave front in circles, as shown in Fig. 295, concentric with O, the *pole* of the wave with respect to P. The areas between successive circles are called ***half-period zones.*** The student may easily find by calculation that these areas are approximately equal. It is evident that the disturbances originating in all points in a circle about O will reach P at the same time, and that the average phases of the resultant effects at P of successive zones will differ by half a period. Although the areas of the zones are practically the same, the amplitude produced by each at P slowly diminishes as its radius increases, on account of increasing obliquity and distance with respect to P. The total amplitude produced by the wave at P is, therefore, the algebraic sum of a series of terms slowly diminishing in magnitude and alternating in sign (direction of displacement). If a_1, a_2, a_3, etc., are the amplitudes at P due to the central area and successive zones, and A the resultant amplitude,

$$\begin{aligned} A &= a_1 - a_2 + a_3 - a_4 + a_5 \cdots \pm a_n \\ &= \tfrac{1}{2}a_1 + \tfrac{1}{2}(a_1 - 2a_2 + a_3) + \tfrac{1}{2}(a_3 - 2a_4 + a_5) \cdots \pm \tfrac{1}{2}a_n \end{aligned}$$

As the successive terms differ very slightly from each other and diminish in accordance with a regular law, the quantities in parentheses and a_n may each be placed equal to zero. Therefore,

$$A = \tfrac{1}{2}a_1$$

or the amplitude at P due to the whole wave is one half and the intensity one fourth that due to the central element if it alone were effective. If the whole wave except the central element is covered the illumination at P will be actually increased, the amplitude in that case being a_1. If all but the two central elements are covered the effect at P is $A = a_1 - a_2 = 0$ nearly. If three elements are uncovered, $A = a_1 - a_2 + a_3 = a_1$ nearly. These

conclusions are easily verified by experiment. If small circular openings of different sizes be placed in a pencil of light diverging from a pinhole, maxima and minima will be found in the centers of the bright areas projected on a screen through the openings (Fig. 296). These holes decrease regularly in size from 1 to 9. If

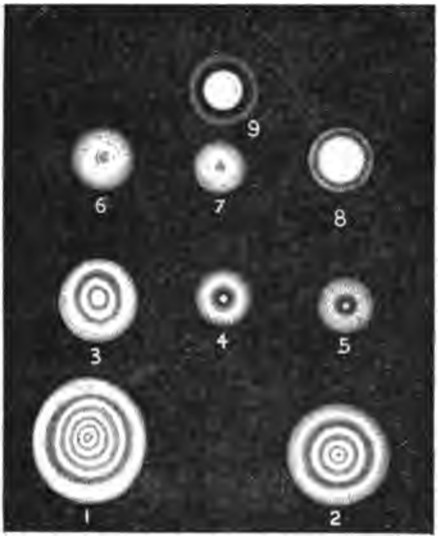


Fig. 296.

the screen be moved (thus changing the number of effective half-period elements subtended by the holes at the screen) maxima change to minima and *vice versa*, or if white light is used the bright spot at the center changes color. The central spot is surrounded by a series of colored bands of similar origin, but not so easily explained by elementary methods. If a hole is smaller than the first two half-period elements, there are no maxima and minima within the illuminated area on the screen, as there can be no possible discordance of phase in the wavelets coming through the hole, and consequently a diffuse circular patch of light is cast on the screen, which increases in size as the opening is made smaller.

If a small disc be placed in the path of the light, so as to cover

a few half period elements as viewed from P—say three—the amplitude of P will be $A = a_4 - a_5 + a_6 - a_7 \cdots = \frac{1}{2}a_4$. A bright spot will therefore be seen at the center of the shadow, nearly as intense as though the disc were removed. At adjacent points of the axis there will be discordance of phase between the disturbances coming around the edge of the disc, resulting in destructive interference.

This remarkable result was first deduced by Poisson as a necessary consequence of the wave theory, and was considered by him as a *reductio ad absurdum* discrediting that theory; but later Arago and Fresnel verified this deduction experimentally. It is easy to repeat this experiment by mounting a perfectly circular disc several millimeters in diameter on a piece of plate glass and placing it in the pencil of sunlight from a small pinhole opening several meters away. The bright spot in the center of the shadow may then be seen on a screen a few meters beyond the disc. A reproduction of a photograph of this effect is shown in Fig. 297.

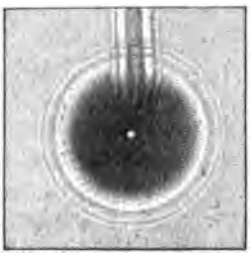


FIG. 297.

476. Straight Edge. If a wave AB from a narrow slit S passes over a straight edge, OD, parallel to the slit, and if the wave be divided into half-period elements with respect to points P, Q, etc., on a screen it may be seen from Fig. 286 that at the points P_0, P_1, P_2, P_3, etc., in the geometrical shadow, each being $\frac{1}{2}\lambda$ further from O than the preceding point, the amplitudes are

$$A_0 = a_1 - a_2 + a_3 - a_4 \cdots = \tfrac{1}{2}a_1$$
$$A_1 = a_2 - a_3 + a_4 - a_5 \cdots = \tfrac{1}{2}a_2$$
$$A_2 = a_3 - a_4 + a_5 - a_6 \cdots = \tfrac{1}{2}a_3$$
$$A_3 = a_4 - a_5 + a_6 - a_7 \cdots = \tfrac{1}{2}a_4, \text{ etc.}$$

Since these terms diminish in magnitude, it appears that the light enters the geometrical shadow, but rapidly fades away. Since the half-period elements of red are larger than those of violet, red penetrates further into the shadow than the other colors, violet the least. This accounts for the spectrum observed in looking at a light over the edge of an obstacle. A part of this effect is, however, due to irradiation (§ 482).

The amplitudes at the points Q_1, Q_2, Q_3, etc., outside the geometrical shadow are, roughly, since some of the elements below the polar lines SQ_1, etc., are effective, at these points,

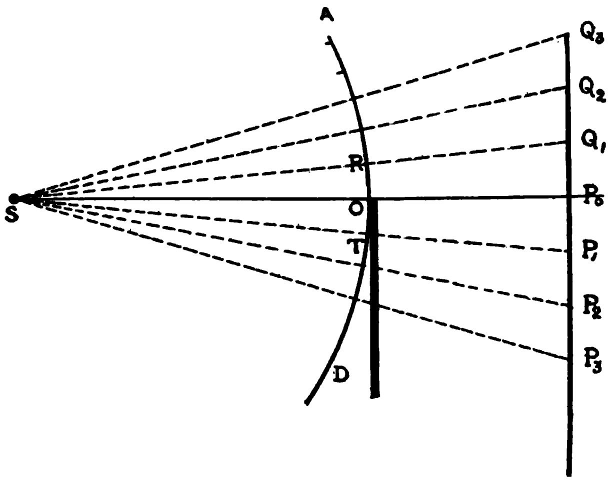


FIG. 298.

$$A_1' = 2a_1 - a_2 + a_3 \cdots = \tfrac{3}{2}a_1 \text{ (maximum)}$$
$$A_2' = a_1 - a_2 + (a_1 - a_2 + a_3 \cdots) = \tfrac{1}{2}a_1 \text{ (minimum)}$$
$$A_3' = a_1 - a_2 + a_3 + (a_1 - a_2 + a_3 \cdots) = a_3 + \tfrac{1}{2}a_1 \text{ (maximum)}$$

Therefore a series of colored bands are seen just outside the geometrical shadow. It must be understood that a different set of half-period elements

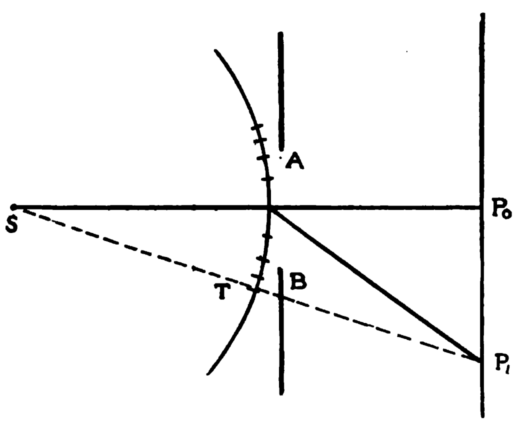


FIG. 299.

must be constructed for each point P, Q; R being the pole of Q_1, T the pole of P_1, etc.

477. Narrow Slit. If two straight edges are opposed so as to form a narrow slit of width AB (Fig. 299) there will be a bright band at P_0 if only the two central half-period elements of an incident wave are exposed. If two on each side are exposed the effect at P_0 is

$$A = 2a_1 - 2a_2 \text{ (nearly zero)}$$

If three half-period elements on each side are exposed

$$A = 2a_1 - 2a_2 + 2a_3 \text{ (maximum)}$$

Therefore there will be successive maxima and minima at P_0 as the slit is widened. If the slit subtends two or any even number of half-period elements as viewed from P_1, a point off the axis (T being its pole), they will neutralize each other in pairs; if it subtends an odd number of such elements, there will be destructive interference between pairs, leaving the odd one effective, consequently there will be a series of maxima and minima on each side of the axis.

It is easy to determine the width of the bands. If $AP_1 - BP_1 = \lambda$ (Fig. 300) we may consider the effects at P_1 of AO and OB to be nearly the same numerically, but to differ in average phase by half a wave length. The two cancel each other. At P_2, where $AP_2 - BP_2 = \frac{3}{2}\lambda$, we may imagine the slit divided into three nearly equal strips, which contribute effects at P alternating in phase. Two cancel each other, leaving the third effective. If D is the distance of the screen from the slit, the width of the central maximum is $2P_0P_1$, and

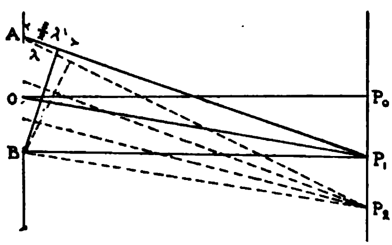


FIG. 300.

$$P_0P_1 : D = \lambda : AB$$

Therefore

$$2P_0P_1 = \frac{2D\lambda}{AB}$$

The other bands are of half this width, or $D \cdot \lambda / AB$. The width of all the bands is, therefore, inversely proportional to the width of the slit. The central maximum is white, the others are narrow spectra bordered by violet on the inside, red on the outside. (Since P_0P_1 for violet is less than the corresponding distance for red.) These effects may be observed by allowing light from a narrow slit to pass through a second adjustable slit and fall on a screen, or more simply by looking through a narrow slit or the space between two fingers at a distant light.

Within and close outside the shadow of a wire or needle cast by a linear source similar fringes are observed. (See Fig. 301, showing shadows of needles of different sizes.)

478. Resolving Power. If light from a narrow slit passes through another slit to a screen the central maximum may be regarded as an image of the first slit (corresponding to a pin-hole image). The wider the second

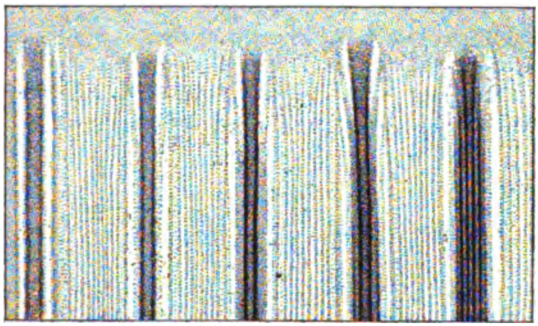


FIG. 301.

slit is opened (up to the point where diffraction effects cease) the narrower and sharper this image will be. Similar considerations apply to light from point sources through circular openings. If we look through a small pin-hole at a distant light it will appear much larger than when viewed with the naked eye. The filament of a lamp appears thicker when seen through a narrow slit. If an image is formed by a lens or mirror the same conditions hold as for a narrow slit, the lens or mirror preserving the uniformity of phase of the whole wave with respect to the focus that exists for a

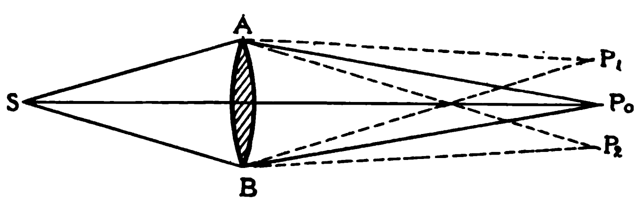


FIG. 302.

narrow slit with respect to its central maximum. Consider the image of a narrow source S (Fig. 302). At P_1 and P_2 on each side of P_0 there will be a minimum if $AP_2 - BP_2 = \lambda = BP_1 - AP_1$ in which case the disturbances from the two halves of the lens reach P_1 and P_2 in opposite phases and cancel each other. The width of the image, which is merely a diffraction

maximum, is, therefore, P_1P_2. If the source is very small such effects arise, and the central maximum will be of the shape of the source, but differently oriented, for any particular dimension of the image will be inversely proportional to the same dimension in the source, as appears from the relation (§ 477).

$$P_1P_2 = 2P_0P_1 = 2D\lambda/AB$$

Observation shows that two diffraction maxima cannot be clearly separated if they are closer than the distance from a maximum to the adjacent minimum. If the image of S_2 for example lies at P_1 it can barely be seen as separate from P_0. The corresponding intensity curves are shown in

FIG. 303.

Fig. 303. The two images will overlap if the angle subtended by the objects at the lens is less than $\alpha = P_0P_1/D = \lambda/AB$. This is called the angle of minimum resolution.

Such conditions apply only to small sources or objects. If the source is large each point will have a large diffraction maximum at the focus. These maxima will overlap and blot out diffraction effects except at the boundaries of the image.

Stars are practically point sources of light. Their images as formed by a telescope with a small objective appear much larger than when formed with a large telescope, the diameters of the central maxima being inversely proportional to the diameter AB of the lens. The image of a double star formed by a small telescope may be one large blur, while that formed by a large telescope consists of two distinct points of light. The ability to separate the images of two small adjacent sources is called resolving power, and as shown above it is directly proportional to the diameter of the lens, mirror, or prism forming the image—or to the cross-section of the effective beam of light, if it does not cover the above.

If we had larger eyes we could see much finer details than we now do. Conversely, it is physically impossible for small insects to see details

clearly. To them an incandescent filament must appear as it does to us when we look at it through a very small pin-hole.

479. Diffraction Grating. If there are a number of narrow and equidistant parallel openings in a screen, each pair of openings will produce effects similar to those observed in Young's double slit experiment. If a lens is placed in front of such a diffraction grating, as it is called (Fig. 304), the same path difference will

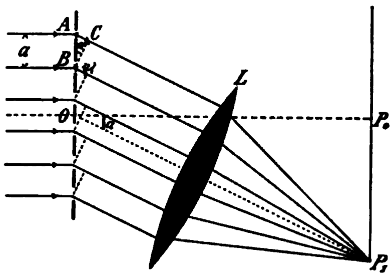


FIG. 304.

exist between any pair of adjacent parallel rays. If $a = AB$ is the distance between openings and if the angle between OP_0 and OP_1 is α, the common difference of path is $AC = a \sin \alpha$, and the condition that there shall be a maximum at P_1 for the wave length λ is

$$a \sin \alpha = n\lambda$$

The central maximum P_0 is white, as the condition for reënforcement at that point ($n = 0$) is the same for all colors. The other maxima are drawn out into spectra on each side of the axis, as α varies with the wave length. The value of the ordinal number n determines the *order* of the spectrum. If α is a small angle the distances between points in the spectra are nearly proportional to the differences of the corresponding wave lengths, so that the spectra formed by gratings are said to be *normal,* as contrasted to those due to prisms, in which there is no simple law of distribution. The grating spectrum is rational, that is, all gratings give spectra which are alike in their distribution of colors, although they may differ in length. The lengths of the spectra increase directly as the order of the spectrum, so that those beyond the first overlap, and they also rapidly diminish in intensity.

The effect due to a grating is precisely the same, so far as position of maxima is concerned, as that due to two slits with the same interval between them. The intensities of the grating spectra, however, are far greater, the amplitude of vibration being in proportion to the number of openings. The resolving power of a grating is also greater. The width of a maximum in the interference bands given by two slits is (§ 416) $w = D\lambda/a$. The width of the maxima given by a grating having N openings is $w = D\lambda/Na$, since Na is the aperture of the grating, so that this width is inversely as the breadth of the grating (§ 477). Diffraction gratings are generally used for measurements of wave-lengths (§ 492).

The first gratings and the first wave length determinations were made by Fraunhofer about 1821. His gratings were constructed of fine wire, or by ruling lines on a smoked glass plate. Later gratings were made by Saxton in America (1843) and by Norbert in Pomerania (1863), by ruling lines on a glass plate with a diamond point. Rutherford, of New York (1863), improved the process of construction, and also devised reflection gratings, ruled on mirrors of speculum metal. The bright reflecting strips between rulings act like illuminated slits. Rowland, of Johns Hopkins University, about 1883 further improved the process by the construction of a nearly perfect screw, with which the carriage holding the mirror could be displaced by equal intervals parallel to itself. Rowland's ruling machines are entirely automatic, so that they may be left to do their work in a constant temperature room without the disturbing influence of the presence of any person. As many as 110,000 lines to the inch may be ruled by these machines, but the usual number is about 15,000.

Cheap copies of a grating may be made by flowing a film of collodion on them, which may be stripped off when dry and mounted between glass plates.

A grating with crossed lines gives a beautiful series of crossed spectra. This effect may be observed by looking through a handkerchief or umbrella top at a distant light. Brilliant diffraction effects are also obtained by looking at a source through a cobweb or feather, or from the light reflected from mother of pearl. In the latter case the effect is due to striations, as may be proved by transferring the effects to wax by pressure.

Corona.—Sometimes a rainbow-colored circle, with violet inside, may be seen around the moon or in looking at a light through fog or dust. It may be artificially produced by sprinkling lycopodium powder over a glass plate and looking through it at a candle flame, the plate being at some distance from the eye. Young showed that the corona, as it is called, is due to diffraction by small particles of the same size but not necessarily uniformly distributed. Consider a beam of light passing in the direction

OP (Fig. 305), the observer's eye being at *P*. It is evident that light passing by the particle *B* will be subject to diffraction. The first maximum for red may be at an angle α with the direction of the light, that for violet at the smaller angle β. All the particles lying on lines at an angle α with the axis will diffract red light to the eye at *P*, while those on a line at angle β with *OP* will send violet to the eye. Diffraction rings will thus be produced, their angular magnitude being evidently inversely proportional to the diameter of the particle, just as the distance between grating maxima is inversely proportional to the distance between lines. From the angular magnitude of the rings the diameters of the particles causing them may be calculated.

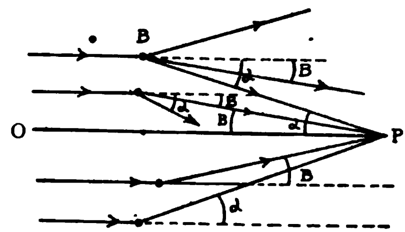


FIG. 305.

OPTICAL INSTRUMENTS AND MEASUREMENTS.

480. The Eye is an essential part of any optical combination. Like a photographic camera, it is a closed chamber into which light can enter only through the lens. As the camera lens throws an image on a sensitive photographic plate which excites the silver grains, the lens of the eye forms a picture on the mat of sensitive nerve endings covering the retina. The amount of light entering the camera is regulated by an "iris" diaphragm of adjustable size; similarly the amount of light entering the eye is controlled by the size of the pupil, which automatically changes in diameter between the limits of about 2 and 5 mm. The parts of the eye are shown in Fig. 306. *S* is the sclerotic membrane, the outer enclosure of the eye. *C* is the cornea, a strong transparent membrane. *I* is the iris, the colored part of the eye, with a central orifice, the pupil, which admits light through the crystalline lens *L*, which focuses images on the retina *R*. The nerve endings covering the retina run together like the strands of a cable into the optic nerve *O*, which conveys stimuli

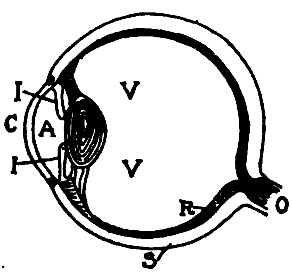


FIG. 306.

to the brain. Muscles attached to the periphery of the lens can by their contraction or relaxation so change its curvature as to enable it to focus either very distant or very near objects on the retina. This process is called accommodation. Two objects are clearly seen separately when the angle between them at the eye is a little less than 1′, or the distance between the retinal images is 0.005 mm. Details are, therefore, more clearly seen as an object is brought nearer, as the angle subtended by it and the size of the retinal image are then larger; but there is a limit to the power of accommodation of the eye, so that usually no object nearer than about 25 cm. can be clearly seen. This is called the distance of most distinct vision. The normal adjustment of the eye when at rest, however, is for "infinity," as may be verified by suddenly raising the eyes when they have been unemployed and looking toward distant objects. They will be in focus.

Between the cornea and the crystalline lens is the aqueous humor, *A*, and between the lens and the retina is the vitreous humor *V*, both transparent fluids with a mean index of refraction equal to 1.336. The lens is built up of transparent horny layers, increasing in density, hardness, and refractive power toward the center. The index of refraction of the outer layer is 1.405; of the next, 1.429, and of the central region 1.454. The average index of refraction is about 1.437. This increase in density toward the axis serves to partly correct spherical aberration, which is also diminished by the iris diaphragm.

Objects such as printed letters can be clearly seen through a pin-hole in a card, even if they are as close as 2 cm. to the eye. This has been attributed to an over-correction of the lens for spherical aberration, so that a narrow pencil passing through the axis of the lens has a very short focus. It is obvious that so much overcorrection would be worse than no correction at all. As a matter of fact, a pin-hole image is formed on the retina, the lens merely sharpening the effect. The fact that the apparent size of the object varies as the card is moved back and forth, the object remaining at rest, shows that the image is due mainly to the pin-hole.

481. Vision. The retina is covered, except over the optic nerve, by a large number of very small fibrous bodies, the "rods" and "cones," nerve endings which are in some way stimulated by

light waves. Over the optic nerve is the "blind spot," so called because if the image falls on this part of the retina it ceases to be visible. By closing one eye and looking steadily with the other at one of two small objects about two inches apart, a distance may be found at which the other object will disappear. Excitation of the optic nerve lasts about one tenth of a second after the stimulus ceases, so that if intermittent stimuli are applied at intervals less than this a steady effect is produced. This is called persistence of vision. The trail of the lighted end of a cigar if it be rapidly moved and the apparent continuity of moving pictures depend on this effect.

Sometimes the normal spheroidal shape of the eye is altered so that the curvatures are not the same in different planes. Light from a point will then pass through the eye as an astigmatic pencil with two focal lines instead of a point image (§ 461). Horizontal and vertical lines at the same distance cannot be simultaneously brought into focus. Such eyes are said to be astigmatic. Other defects arise from change of curvature or from loss of the power of accommodation. If eyes are short sighted, the principal focus falls short of the retina, and distant objects cannot be clearly seen. If they are long sighted, the principal focus lies on or near the retina, and images of near objects cannot be formed on the retina. For the first defect concave spectacles are the remedy; for the second they must be convex.

In normal eyes the nerve endings on which fall corresponding points of the two retinal images lead to the same nerve centers, so that the two pictures are exactly superimposed and a more intense effect secured than with one eye alone. If one eye-ball be forcibly twisted out of position double images will be seen. A further advantage given by two eyes is that an object is viewed from two slightly different directions, which gives the impression of relief. This principle is applied to the stereoscope, in which two photographs taken from slightly different points of view are viewed by each eye separately. The two images will be superimposed in such a manner that the object appears to stand out in space.

With two eyes it is also easier to estimate distances than with one. There is an angle between the two lines of sight to the object, which the brain unconsciously estimates. In general the sizes of objects are inferred from their angular magnitudes and estimates of their distance based on ex-

perience, or by comparison with adjacent objects, such as trees and houses, the sizes of which are approximately known. Such estimates are influenced by the clearness with which details are seen. In places where the atmosphere is unusually clear, as in Arizona, this leads to the underestimation of distance. Conversely, objects seen in a fog appear to be more distant than they are, owing to the indistinctness of their details. The angle subtended by them, however, corresponds to the actual distance, hence they loom larger than they are.

482. Irradiation is the apparent increase in size of objects as they become brighter. The crescent of the new moon, for example, looks larger than the remainder of the disc, the "old moon," which is illuminated by the earth alone. The filament of an incandescent lamp appears to increase in size as it passes from ordinary temperatures through red and white heat. This effect was long supposed to be due to the spreading of the retinal image on account of stimulation of nerves outside of its boundaries, in much the same way that an overexposed photographic image is affected. It is now believed by some that the effect is due merely to spherical aberration of the eye, which becomes more noticeable as the intensity of the source increases.

483. The Simple Microscope or magnifying glass is a single convex lens through which objects at or within the principal focus of the lens are viewed. As shown in Fig. 307, an enlarged virtual image $A'B'$ is formed subtending at the lens the same angle α as the object AB. The linear size of this image is determined from the relation

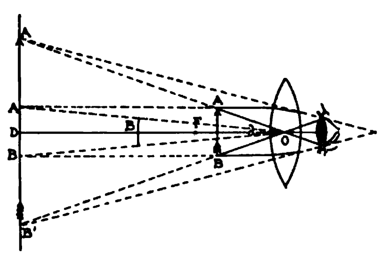


FIG. 307.

$$I = (v/u)O$$

As the normal adjustment of the eye is for infinity, the object is usually at or very near the principal focus. In no case can the image be clearly seen when nearer than the limit of distinct vision. The actual linear magnitude of the image counts for little; the size of the retinal image depends on the angle subtended at the eye, and if the latter is very near the lens this angle is substantially

that subtended from the lens. The lens simply increases the power of accommodation of the eye, so that the object may be brought nearer and thus subtend a greater angle. With the unaided eye, the greatest detail is observed at the distance of most distinct vision (25 cm.) where it subtends the angle β (Fig. 307). With the lens, the object is brought nearer, approximately to the principal focus, and the angle subtended by it increases from β to α. The magnification M of the retinal image is, therefore, α/β. If f is the focal length of the lens, d the limit of distinct vision

$$AB = 2d \cdot \tan \beta/2 = 2f \tan \alpha/2$$

Therefore, if these angles are small

$$M = \alpha/\beta = d/f$$

Power of Lenses. The magnifying power of a lens is, as shown above, inversely proportional to the principal focal length, hence $1/f$ is a measure of its power. The practical unit of lens power is that of a lens with a focal length of one meter. This unit is called a *diopter* or *dioptric*. The power of converging lenses is considered positive, that of diverging lenses negative. The relation deduced in § 462 shows that the power of a number of lenses in contact is the algebraic sum of their individual powers.

484. Eyepieces. The part next the eye of an optical train of lenses, such as those of telescopes and compound microscopes, usually consists of some form of simple microscope known as an eye piece. With a single lens, much of the light from the real image formed by the objective O, which is usually viewed through the eyepiece, would be lost. In order to avoid this, light is gathered in toward the axis by a second lens, called the field lens F (Fig. 308). Nearly all the light would pass by the edge of the eye lens E if F were absent. It may be shown that a combination of two lenses of the same kind of glass is nearly achromatic if they are placed at a distance from each other

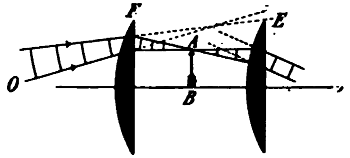


FIG. 308.

$d = (f_1 + f_2)/2$. This property is utilized in most eyepieces which consist of a field lens and eye lens.

In Huyghens' eyepiece $f_1 = 3f_2$ (Fig. 308). Hence $d = 2f_2$ and if the image due to the objective and the first lens is formed half way between the lenses the emergent light will be parallel and a virtual image formed at infinity. If a cross thread is used, it must be placed at AB. The lenses are convex toward the incident light and of such curvature as to reduce the spherical aberration to the minimum.

In the Ramsden eyepiece (Fig. 309) $f_1 = f_2$. If the lenses are placed apart at the distance $(f_1 + f_2)/2$, dust particles on the field lens would be visible through the second. In order to avoid this, the lenses are usually placed at a distance of $2f/3$. The principal focal point of the combination is at a distance $f_1/4$ in front of the first lens. The object, or the real image due to the objective, is at this point, and the final virtual image is at infinity. The chromatic aberration is small, and the spherical aberration is reduced by using plano-convex lenses with convex surfaces facing each other.

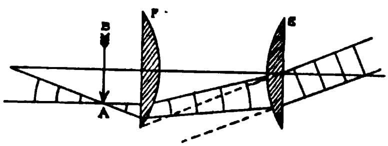


FIG. 309.

In all these eyepieces the emergent red and violet rays are nearly parallel, hence the virtual images formed by the different colors subtend very nearly the same angle at the eye, and are, therefore, of the same size, but not quite equally sharply focused on the retina.

485. Compound Microscope. In order to extend the limit of magnification beyond the point obtainable with a simple microscope, a combination of lenses is used. An enlarged real image $A'B'$ (Fig. 310) is formed by an object lens or train of lenses, and this image is further enlarged by an eyepiece, such as that of Huy-

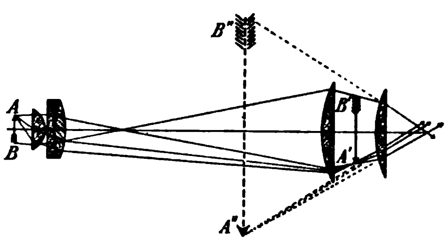


FIG. 310.

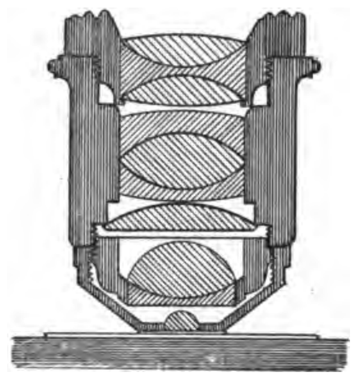


FIG. 311.

ghens, used as a simple microscope, which gives a virtual image $A''B''$. The front lens of the objective train is usually of the hemispherical form described in § 448, which has a great angular aperture, with very little spherical aberration. There are in addition a number of other lenses of different shapes and kinds of glass, so combined as to reduce spherical and chromatic aberration to a minimum and to give a plane focal surface. A typical combination is shown in Fig. 311.

The magnifying power of the objective, of focal length f_1, is

$$M_1 = I_1/O = v_1/u_1$$

That of the eyepiece is, as shown in § 474,

$$M_2 = I_2/I_1 = d/f_2$$

where d is the minimum distance of distinct vision. The magnification due to the combination is

$$M = M_1 M_2 = I_2/O = Ld/f_1 f_2 \text{ approximately}$$

where L is the distance between the objective and the eyepiece.

The minimum distance between two small objects A and B seen through a microscope which will permit of clear separation of their diffraction images is obtained by a slight modification of the expression found for the minimum angle of resolution, $\alpha = \lambda/A$ (§ 478). The minimum value which d can have is thus found to be $\lambda/2$, when the object is at the surface of the lens. Since this distance is proportional to the wave length, details which may be clearly seen when the object is illuminated by blue light will be indistinct when red light is used.

486. Astronomical Telescope. The object glass of a telescope forms a real and of course greatly reduced image $A'B'$ of a dis-

tant object (Fig. 312). The object and its image subtend the same angle α at the objective, and the object subtends practically the same angle α if viewed directly by the eye. If, however, the eye views the image formed by the objective at the distance of most distinct

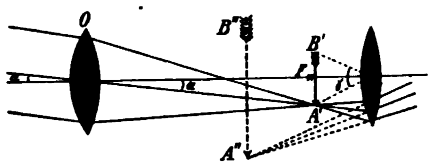


FIG. 312.

vision, this image will subtend an angle β which is larger than α, and the apparent magnification is $M_1 = \beta/\alpha$. When this image is viewed through an eyepiece, there is further magnification, the image subtending the larger angle γ. The magnification due to the combination is

$$M = M_1 M_2 = \frac{\beta}{\alpha} \times \frac{\gamma}{\beta} = \frac{\gamma}{\alpha} = \frac{f_1}{f_2}$$

The limiting angle of resolution between two linear sources is proportional to λ/A, where A is the diameter of the objective (§ 478).

For astronomical purposes there is no disadvantage arising from the fact that an inverted image is formed by a telescope, but when the instrument is to be used for terrestrial purposes it is necessary to add an additional lens or pair of lenses to reinvert the real image formed by the objective. This adds inconveniently to the length of the tube. If the image is inverted by reflection from a combination of prisms the length may be diminished, but for most purposes where only small magnification is required the form of telescope devised by Galileo is most convenient.

487. Dutch or Galilean Telescope. This type is used for opera glasses and for marine glasses. As shown in Fig. 313 an erect virtual image is formed, the magnification being $M = f_1/f_2$ (§ 486). The tube has a length approximately equal to the difference between the focal lengths of the objective and the eyepiece, while in the ordinary telescope the length is the sum of these distances.

488. Reflecting Telescope. The objective lens may be replaced by a large concave mirror. In this way chromatic aberration may be entirely avoided, but spherical aberration is more trouble-

some than with refractors. As the real image is formed along the axis of the mirror and in the path of the incident light, special devices are necessary in order to view it. In the Newtonian telescope the image is reflected to one side by a small right-

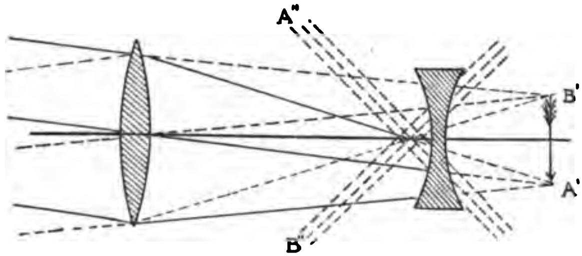


FIG. 313.

angled prism which cuts off very little light, and is viewed by an eyepiece in the side of the tube. Herschel tipped the mirror slightly so that the image was formed at the edge of the open end of the tube, at which point the eyepiece was fixed. In other forms a small mirror in the axis reflects the image back into an eyepiece set in the center of the objective itself, so that it can be viewed from behind.

The first telescope was probably made by Lippershey, in Holland about 1608, with a concave lens for eyepiece. Rumors of this discovery reached Galileo next year, and he immediately devised a telescope having a plano-convex objective and plano-concave eyepiece. The discovery of Jupiter's satellites and Saturn's rings soon followed. This subversion of the accepted order of things greatly scandalized the orthodox. Kepler introduced the convex eyepiece about 1611. In order to diminish the spherical aberration, Huyghens and others used objectives of very great focal length. The objective and the eyepiece were mounted at opposite ends of rods of great length.

Newton devised the reflecting telescope in order to avoid chromatic aberration. After Dolland's discovery that chromatic aberration may be corrected, the refracting telescope again came into favor.

Some very large reflecting telescopes have been made, the largest being that built by Lord Rosse in 1845, which had an objective six feet in diameter. One under constructon for the Mount Wilson Solar Observatory will be 100 inches in diameter. The largest refracting objective is that of the Yerkes Observatory of the University of Chicago, which has an objective forty inches in diameter and nearly eighty feet focal length.

The use of a single lens as "burning glass" and magnifier is of unknown antiquity. The invention of the compound microscope is ascribed to various persons, to Galileo among others. At any rate he appears to have been led independently to its construction by his success with the telescope. The instrument has reached its present excellence largely by reason of Amici's discovery of the hemispherical lens and of the effects of immersion, and of Abbe's improvements in the manufacture of glass of suitable optical properties, and in the theory and art of lens construction.

489. Photographic Camera. This is a form of camera obscura in which the image formed by a lens falls on a sensitive photographic plate. The requirements demanded for the lens are exacting and in some cases contradictory to each other. It must give images free from spherical and chromatic aberration, and in many cases have great light power and a large field of view. The focal surface must be plane, and the magnification must be the same in all parts of this plane, so that no distortion is produced. The depth of focus must be great, that is, objects at different distances must have images approximately in focus at the same time on the plate. As the film is most sensitive for the shortest

FIG. 314.

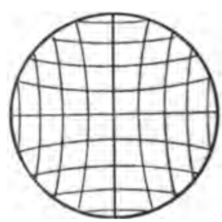


FIG. 315.

waves, the lenses must be corrected for the violet and the yellow, instead of blue and red. A diaphragm with small openings is used in front of the lens, if it is a single achromatic combination, such as is used for landscape work. This reduces spherical aberration and at the same time gives a greater depth of focus (approximating to the principle of the pin-hole camera, in which the focus is nearly independent of the distance). A diaphragm with a single lens results in a distorted image, however, as shown in Fig. 314, which represents the distortion of a quadrilateral network with the diaphragm in front of the lens, and Fig. 315, which gives the effect due to a diaphragm behind the lens. The cause

is readily seen to be due to the difference in deviation of pencils passing through the center and the edge of the lens respectively. If two lenses are used, with the diaphragm at the optical center of the combination, these distortions correct each other.

As shown in Fig. 316, the perfect symmetry of the incident and the transmitted secondary axes AA', BB', CC', etc., with respect to the opening O shows that the distances AB, BC, etc., in the object are in the same ratio

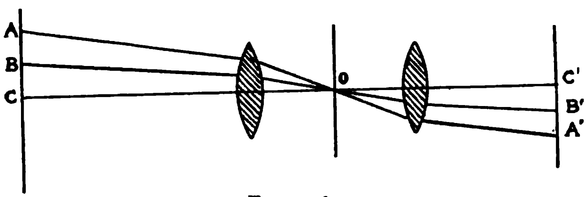


FIG. 316.

as the corresponding distances $A'B'$, $B'C'$, etc., in the image, so that there is no distortion, and if A, B, C are in the same plane, $A'B'C'$, etc., must be in the same plane. Such lenses are called rectilinear or orthoscopic doublets.

The size of the photographic image of a distant object is nearly proportional to the focal length of the lens. It is, however, inconvenient to give a great length to the camera box. This difficulty is avoided by the use of the teleobjective, in which a concave lens L_2 is placed behind the converging lens L (Fig. 317). The divergent effect of this lens gives a virtual focal length equal to PF, while the camera box has the much smaller length LF. A greatly enlarged image is secured, but the field of view is reduced.

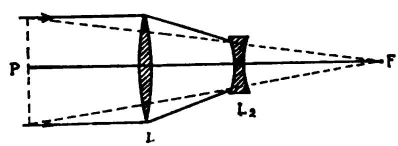


FIG. 317.

490. The Projection Lantern is used to throw an enlarged image $A'B'$ of more or less transparent objects on a screen. The object AB (Fig. 318) is illuminated by a condenser C, consisting of two thick plano-convex lenses, with convex sides facing each other. The sources are usually the electric arc or the calcium light. The focusing lens L is generally of the

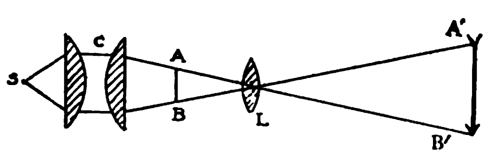


FIG. 318.

photographic doublet type, in order that an undistorted image may be formed on the screen. The object of the condenser is not only to illuminate the object, but also to enlarge the field beyond the limit which would otherwise be set by the cross-section of the focusing lens.

491. The Spectroscope is an instrument designed to analyze complex radiations by prismatic dispersion (§ 447) or by the diffraction grating. In order to secure as complete separation of the colors as possible, or a "pure" spectrum, a narrow slit must be used as a source, so that the colored images of the slit will overlap as little as possible. The resolving power must be so great that the diffraction images of the slit or "lines" do not overlap, and this requires large apertures for the lenses and prism or grating (§ 478). The larger the dispersion the more complete the separation of the images. For given dispersion, the length of the spectrum is proportional to the focal length of the observing telescope, but this merely affects the scale of the spectrum, not the resolution of the lines or the clearness of detail.

The plan of the ordinary form of spectroscope is shown in Fig. 319. The essential parts are: A narrow slit, S; a collimating lens C, which converts the wedge of light from the slit into a parallel beam; a prism to disperse the colors; a telescope lens T, with which real images of the slit are produced at the focus of the eyepiece E. If light of an infinite number of colors is emitted by the source, the infinite number of partially overlapping images forms a continuous spectrum; if only a finite number of colors are emitted, there will be a finite number of slit images, giving a discontinuous or *line* spectrum. As shown in § 446, a homocentric pencil (cone) incident on a prism remains homocentric after transmission at the angle of minimum deviation; for all other angles of transmission, it becomes astigmatic, and no true image is produced. The condition of homocentricity cannot be fulfilled for all colors simultaneously except when the incident light is parallel, in which case light of each color emerges in a parallel beam. For this reason

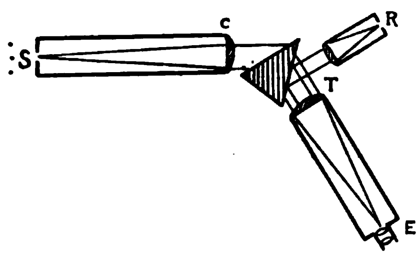


FIG. 319.

the collimator is necessary. This lens must be achromatic. So far as purity of spectrum is concerned, it is evidently unnecessary for the telescope lens to be achromatic, but it is usually corrected, in order that all the colors may be at once in the focus of the eye-piece. Positions in the spectrum may be referred to the image of a scale R reflected from the side of the prism.

This instrument is called a spectrograph when the telescope is replaced by a camera for photographing the spectrum, and spectrometer when provided with a graduated circle for measuring the angular deviation of the light. The direct-vision spectroscope, usually made in small sizes for pocket use, has a combination of crown and flint glass prisms, as shown in Fig. 320. The mean deviation is zero, but there is some residual dispersion which gives a short spectrum (§ 451).

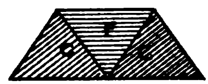


FIG. 320.

A plane diffraction grating may replace the prism of a spectroscope, or spectrometer. With the latter the angular deviations of the diffraction maxima may be measured and the wave lengths determined by the relation deduced in § 479.

492. The Concave Grating was Rowland's greatest contribution to spectroscopy. The lines are ruled at equal distances on the surface of a concave mirror of speculum metal which focuses as well as diffracts the light. If R is the radius of curvature of the mirror (Fig. 321) and if the slit is at any point on the circumference of a circle having the radius as its diameter, it is found that the spectra of all orders are in focus, along the circumference of this same circle, so that no lenses are necessary. Usually the grating G is mounted at one end and the eyepiece, or camera, E at the other end of a beam R equal in length to the radius of the grating. This beam has a swivel truck under each end, which travel on tracks at right angles to each other, with the slit at the intersection S. The distance SE between the slit and the eye-piece are proportional to $a \sin \alpha = n\lambda$, and therefore are pro-

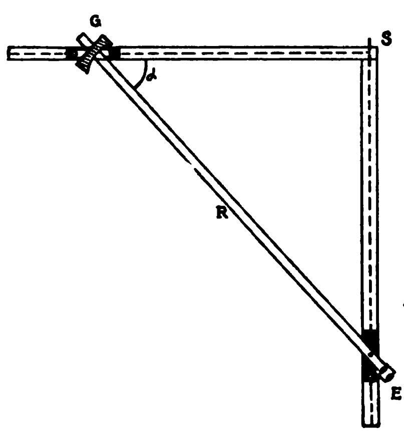


FIG. 321.

portional to the wave lengths of the part of the spectrum in the field of view.

493. Michelson's Interferometer. The surface of the glass plate P_1 (Fig. 322) is "half silvered," that is, the silver film is of such thickness that about one half of the incident light is reflected. Light from the point S of an extended source falls on this surface at A and is in part transmitted to the mirror M_1, in part reflected to the mirror M_2. From these mirrors it will be reflected, retrace its course, and some will finally reach the eye at E. If M_1 and M_2 are at the same optical distance from S, and if they are parallel, the light will appear to come from two exactly superimposed images of the source and there will be no interference. The plate P_2 is introduced merely to give the ray SAM_1 the same path in glass as the ray SAM_2, so that the optical and the geometrical paths will be the same. Now if the mirror M_1 be displaced to M_1' a distance d the two images of the source will be displaced a distance $2d$ but will remain parallel (Fig. 323). The difference of path of the rays coming from two corresponding points S_1 and S_2 of these images to the eye at E is $2d \cos i$. Along the axis SO the difference of path is $2d$ and around this are circular fringes corresponding to the loci of $\cos i =$ constant. (§ 471.) If the mirrors are

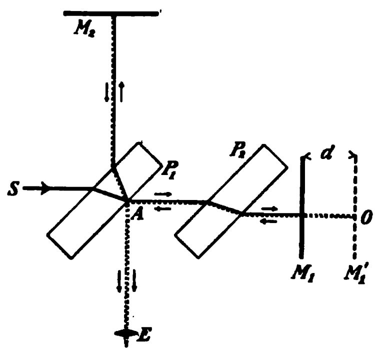


FIG. 322.

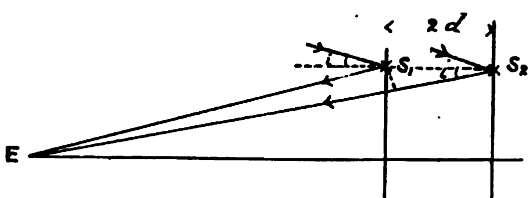


FIG. 323.

not quite parallel, the images will be inclined at a small angle, and the effects produced will be similar to those due to a wedge-shaped film. The fringes will be parallel and straight.

By slowly displacing one mirror and counting the number of fringes N displaced across the field, the distance through which the mirror has moved may be determined from the relation

$$2d = N\lambda$$

Michelson has by this method measured the length of the standard meter in Paris in terms of the wave lengths of several of the spectral lines of cadmium. The wave lengths of spectral lines are probably the most permanent and unchangeable standards of length which can be obtained.

If the light has only two components of slightly different wave lengths, as in the case of the *D* lines of sodium, and if the mirror M_1 be displaced from the position where the optical paths AM_1 and AM_2 are equal, the two sets of fringes due to the two components will gradually get out of step. When the maxima of one set are superimposed on the minima of the other, the field will be nearly uniformly illuminated, and the visibility of the fringes is small. If the mirror be further displaced, the fringes will once more get into step, their maxima and minima will coincide, and the visibility will be great.

When the visibility first reaches a minimum

$$2d = n\lambda_1 = \frac{2n+1}{2}\lambda_2 \qquad \text{or} \qquad \lambda_1 - \lambda_2 = \frac{\lambda_2}{2n} = \frac{\lambda_1\lambda_2}{4d} = \frac{\lambda^2}{4d}$$

In this way, if the average wave length λ of the two lines is approximately known, very accurate measurements of the difference of their wave lengths may be made, and two lines which are so close that they cannot be resolved by a grating may be separated if d is made sufficiently large.

494. Fabry and Perot Interferometer. In the Michelson instrument only two interfering pencils are effective. The interference bands are, therefore. broad and diffuse. like those produced by the interference of light from two slits. Fabry and Perot devised an interferometer composed of two half-silvered parallel glass surfaces *AB* and *CD* (Fig. 324). If light from an extended source is incident normally or nearly so a large number of rays arising from multiple reflection will be in a condition to interfere, and the fringes may be seen in the transmitted light. The resolving power of such an arrangement is proportional to the number of components, as in the case of a grating, so that the bands are very narrow, like the spectral lines due to the latter. The uses of the instrument are the same as those of the Michelson interferometer, but it is more powerful. It is not necessary to restrict observations to positions of the fringes giving maximum

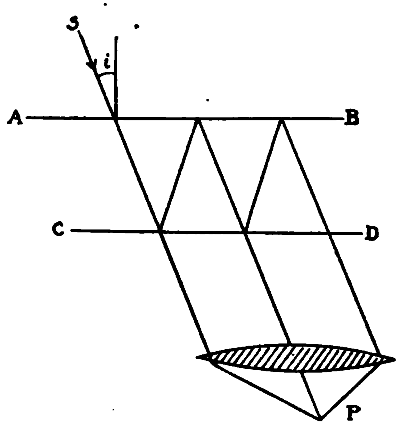


FIG. 324.

and minimum visibility, because the fringes are so narrow that when light of two different wave lengths is used, a separation of one set from the other as small as one twentieth of the distance between fringes is easily detected, or the resolving power is some ten times as great as with the Michelson interferometer.

With these and similar forms of interferometers differences of path between interfering beams of several hundred thousand wave lengths have been reached.

495. In the **Measurement of Refractive Index** of solids in the prismatic form the relation

$$n = \frac{\sin \frac{A + D}{2}}{\sin \frac{A}{2}}$$

is used. The angles of the prism and of medium deviation are measured on a spectrometer. The same method may be applied to liquids if they are contained in hollow prisms with glass slides.

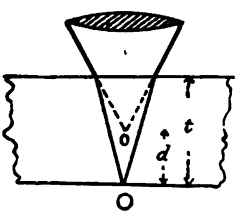


FIG. 325.

The index of refraction of plane parallel plates may be obtained from the relation deduced in § 443. A microscope is focused on a small object on a table, such as a pencil mark O (Fig. 325). When the plate is placed over the mark it will be necessary to raise the microscope a distance d to bring the virtual image O' into focus. The apparent depth of the object below the surface is $t' = t/n$, and $d = t - t' = t - t/n$. Hence $n = t/(t - d)$.

The index of thin films may be found with the Michelson interferometer. If a film of thickness t be introduced in the path of one ray the change of optical path produced is $2(n - 1)t$. (§ 453.) This causes a shift of the fringes equivalent to that due to moving one mirror half this distance.

The index of refraction of a liquid or of a small portion of an opaque object may be determined by measuring the angle of total reflection from its surface when in contact with a more refractive medium and using the relation $\sin k = n/n_1$, where n is the index of the less and n_1 that of the more refractive medium.

496. Lummer-Brodhun Photometer. There are a number of forms of photometer in use besides those described in § 407. One of the most useful of these is that of Lummer and Brodhun. A cube C is made of two right-angled prisms, as shown in Fig. 326. The hypothenuse surface of the one prism is plane, that of the other convex, with the vertex ground flat. These two surfaces are in close contact. The sources to be compared, S_1 and S_2, are mounted on an optical bench, over the center of which is the white screen W of paper or gypsum. The diffuse illumination from this screen is reflected from the mirrors M_1 and M_2 through the prism faces AB and CD. Light from S_1 and S_2 is transmitted without loss through the area of contact of the two prisms, and is totally reflected from the air film between those parts of the hypothenuse surfaces which are not in contact. If a telescope is focused on the region of contact through the side CD, light will enter it from S_1 by transmission and from S_2 by reflection. The field will be uniformly illuminated if the two sides of the screen W are equally illuminated; otherwise the area of contact will appear brighter or darker than the surrounding part of the field. The arrangement is somewhat similar to the Bunsen grease-spot photometer. There are no corrections to be made, because the system is symmetrical with respect to both sources, and each component of light is subject to exactly the same losses due to reflection and absorption.

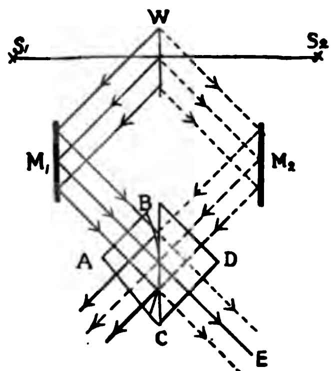


FIG. 326.

497. Standards of Luminosity. For accurate and easily reproducible comparisons of photometric measurements constant and easily attainable standards are necessary. So far no absolutely reliable standard source has been found. For ordinary purposes the British standard candle is used. These are made of sperm, weigh six to the pound, and are normally supposed to burn 120

grains per hour. In actual practice there are great deviations from uniformity.

Other standards in use are the Methven screen and the Hefner-Alteneck lamp. The former is a gas flame from an Argand burner, the light from which passes through a rectangular opening of definite size. The latter is the flame of a lamp burning amyl acetate. In both, the flame should be kept at a definite and constant height. The light from these sources is constant within a few per cent.

EMISSION OF RADIANT ENERGY.

498. Analysis of Radiation. The methods by which radiation may be analyzed by the dispersion of colors, or waves of different length, have already been described (§ 491). The radiation from all known sources is complex—that is to say, it contains waves of more than one frequency of vibration. The principal types of emission spectra may be observed side by side if the image of a long electric arc is focused on a slit beyond which a prism and lens are placed so that a large spectrum is thrown on the screen. The light coming from the positive carbon forms a brilliant continuous spectrum including all the colors. Next to it is the discontinuous spectrum of the arc proper, the luminous flame, which contains the vapors of carbon, various compounds of carbon, and any metals that may be present as impurities in the electrodes. This spectrum consists of a number of narrow lines due to the metals present, and several groups of bands, each composed of a large number of fine lines so spaced as to produce the effect of the shading of fluted columns in a line drawing—hence they are often referred to as fluted bands (see Plate opp. p. 455). These appear to be due to the vapors of carbon or the compounds of carbon. The bands are especially strong in the violet, and this gives the arc its characteristic violet color. The violet bands appear to be mainly due to cyanogen or some other compound of carbon with nitrogen, as they are very weak if the arc is deprived of nitrogen. Next to this is the spectrum of the negative electrode, which is at a much lower temperature than the positive, in which there is very little blue or violet.

As illustrated by the arc spectrum, there are two general types

of *emission* spectra, the *continuous* and *discontinuous*, and the latter in turn may be divided into *line* and *band* spectra.

499. Invisible Radiations. It was found by William Herschel in 1800 that if a sensitive thermometer is placed in any part of the spectrum of the sun it will show a rise of temperature, this effect increasing in going from violet to red. It does not, however, cease abruptly at the boundary of the visible spectrum, but increases to some distance beyond it, and then gradually diminishes, the observed limit in any case depending on the sensitiveness of the thermometer. Evidently there is radiation which is less refracted than the red, and which by analogy we may conclude has waves of greater length than those of red light. It was shown by Herschel that this "radiant heat" is subject to the same laws of reflection and refraction as light, but their identity was not generally accepted until nearly fifty years later, when Fizeau and Foucault showed that the invisible radiation is capable of producing interference effects, and Melloni proved that it is likewise capable of dispersion and polarization. As the ideas in regard to the nature of heat crystallized it was seen that heat can be associated only with matter, but that the energy of this heat may be partly transformed into the energy of ether waves, and this, if absorbed by matter, will again appear as heat. It thus becomes clear how invisible radiations from a hot body can pass through an ice lens without melting it, and set fire to a piece of paper at the focus. The name *infra-red* is applied to these long-wave radiations.

In 1801 Ritter showed the existence of *ultra-violet* radiations in the solar spectrum, with waves shorter than those of violet, by means of the chemical effect produced on chloride of silver. Later it was found that photographic films are very sensitive to the ultra-violet radiation, which is especially active in its chemical effects. It also excites strong fluorescence (§ 525) in many substances. If a strip of paper moistened in acidulated sulphate of quinine solution is held in the arc spectrum the excited fluorescent light shows the existence of ultra-violet radiation in the spectra of both the positive carbon and the arc proper.

For a long time after the undulatory theory of "radiant heat" was firmly established, and it was made clear that the non-visi-

bility of the infra-red and ultra-violet radiations is due merely to the limitations of the eye, the artificial distinction between "heat," "light," and "actinic" spectra held its ground. The eye will "resonate" to vibrations between certain limits of frequency, the photographic film or fluorescent screen to certain others; but if the receiving surface is blackened the energy of waves of all frequencies is almost completely absorbed, and by the amount of heat developed we may determine the amount of energy in any part of the spectrum.

500. Methods of Detecting Invisible Radiation. Photography is a thoroughly satisfactory method of detecting even the shortest ultra-violet radiations so far discovered. This method cannot be used, however, at the opposite end of the spectrum, as it seems impossible to make any photographic film which is sensitive to the infra-red—it is difficult to make one which will even reach the limit of the visible red. For this reason other less satisfactory methods must be employed, which are usually based on the heating effects produced. For one of the earliest instruments for the detection of infra-red radiation, the thermopile, as well as other more recent modifications, see § 307.

501. Continuous Spectra. It is a familiar experience that as the temperature of a body rises it first reaches a dull red heat, then yellow, and finally a dazzling white. Conversely, if the spectrum of the positive electrode of an arc light is thrown on a screen, and if the current is suddenly cut off, it will be observed that, as the carbon cools, violet, blue, green, and yellow disappear in succession, and finally the red. If a sensitive thermopile is placed far in the infra-red it will be found that sensible radiation is still emitted long after the luminosity has disappeared.

Draper (1847) found that all bodies begin to glow at about the same temperature. The actual temperature in any case depends somewhat on the sensitiveness of the eye, but is not far from 400°.

After the invention of the bolometer and other instruments for examining the details of infra-red spectra progress was rapid. Paschen (1897), Lummer and Pringsheim (1898), and others have shown that Draper's law is approximately true for all colors and temperatures—that is to say, all solids begin to radiate red, or yellow, or violet, or any particular "heat color" in the infra-red at the same temperature. The spectral distribution of energy may

be shown by plotting a curve with wave lengths as abscissæ and ordinates proportional to the galvanometer deflections observed as the bolometer strip passes through the spectrum. Fig. 327 shows a series of such curves obtained by Lummer and Pringsheim for temperatures ranging from 836° to 1377° absolute, the source consisting of a strip of blackened metal electrically heated. Langley has made a similar energy curve for the radiation of

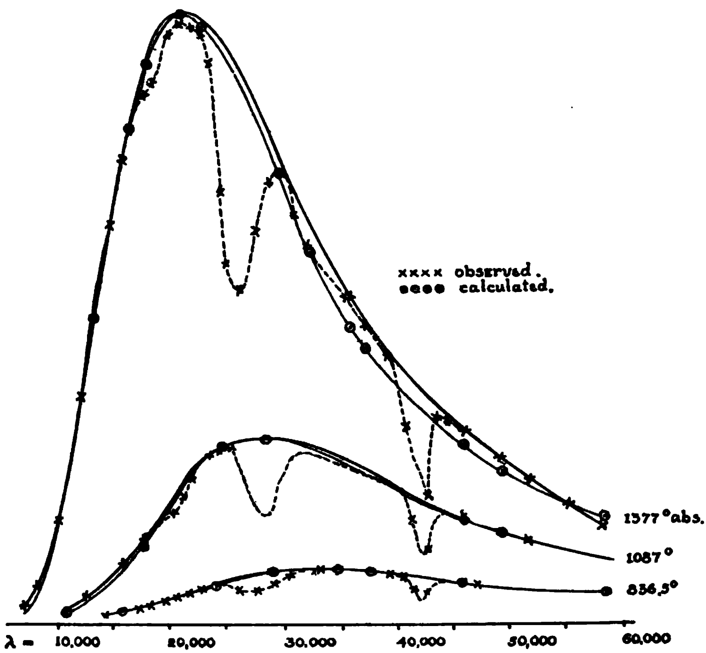


FIG. 327.

copper at a temperature of —2°. The depressions in the curves are due to absorption by carbon dioxide and water vapor. The general character is the same for all solids, but differences in the ordinates may arise from differences in the state of the surface, whether black, or rough, or polished, etc. Investigation shows that there is a very definite relation between the absolute temperature of the source if it is black, or approximately so, and the wave length corresponding to the maximum ordinate of the energy curve,

such that $\lambda_m T = \text{Constant}$ (see § 314). This constant varies slightly from 2814 at 621.2° to 2928 at 1646°, with a mean value of 2879. The unit of wave length is the micron, or .001 mm.

The total energy emitted by an incandescent source is proportional to the area included between the energy curve and the axis of X. That part of the energy which produces luminosity is included between the limits of the visible spectrum. The luminous efficiency is proportional to the ratio between these two areas. Fig. 328 illustrates the relative luminous efficiencies of the positive pole of an arc light, of an incandescent light, and of a piece of red hot carbon. The luminous energy is represented by the shaded part of the area between the curve and the λ-axis. Evidently the luminous efficiency rises very rapidly with the temperature, so that a small increase in the current through an incandescent lamp will greatly increase its brightness, and conversely. The luminous efficiency of the arc is about 10 per cent.; of incandescent lamps, 3 to 5 per cent.; of gas and candle flames 2 or 3 per cent. Luminous vapors, which radiate "selectively," are usually more efficient than solids.

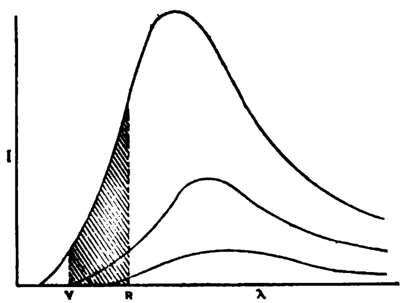


FIG. 328.

502. Law of Radiation. It may be assumed that ether waves are set up by molecular agitation in matter. In solids the molecules are so close that there can be little chance for them to vibrate without constraint with their natural periods. Owing to frequent collisions, a wide range of velocities and vibration frequencies must exist. The forced ether vibrations will have periods corresponding to those of the molecules. The greater part of the radiant energy will be due to the large number of molecules which have velocities in the neighborhood of the mean velocity. There will be relatively few molecules which will have extreme velocities, either large or small, and therefore the longest and shortest ether waves will have relatively a small amount of energy. As the temperature rises there will be a general increase of kinetic energy, many mole-

cules moving faster but none slower that before, so that both the maximum energy and the maximum rate of vibration of the excited ether waves will move toward the violet end of the spectrum. The source will rise to red and finally to white heat. It is thus evident that a spectral intensity curve must be a sort of probability curve based on the distribution of molecular velocities, with its ordinates exaggerated on the side of the violet, as illustrated in the curves of Fig. 328.

By such reasoning Wien (1893) theoretically deduced the relation $\lambda_m T = C$, known as Wien's law (previously suggested by Weber). Later Wien established a general relation between the intensity of radiation corresponding to a given wave length λ, and the absolute temperature of the source, as follows :

$$I_\lambda = \frac{C}{\lambda^a} e^{-\frac{c}{\lambda T}}$$

where e is the base of the natural system of logarithms, T the absolute temperature, and C and c constants. The work of Paschen and others shows that this relation holds within wide limits for bodies which are black or approximately so. For such bodies the value of $a = 5$, $c = 14395$, $C = 1569 \times 10^{11}$. The law $\lambda_m T = \text{constant} = c/5$ may be deduced by differentiating the above expression for a maximum value of I_λ, and Stefan's law by integrating the intensity over the whole spectrum (see § 313).

503. Discontinuous Spectra were noted by a number of observers during the first half of the nineteenth century, and it was at least dimly realized that in some cases the number and position of lines or bands is characteristic of the substances emitting the radiation.

In 1860 Kirchhoff and Bunsen established definitely the law that all gases and vapors give discontinuous spectra, and that these spectra are perfectly characteristic of the substance. They discovered rubidium and cæsium by the application of this principle, which has been fruitful in the discovery of other new elements, notably helium and the rare atmospheric gases in recent times.

At first only the spectra of flames were studied, but it was found by Ångström (1850) that the electric spark between metallic terminals gives lines due both to the electrodes and to the surrounding atmosphere, and Plücker (1858) found that if the electric discharge passes through a gas in a partially exhausted tube (vacuum tube) the luminosity is confined to

the gas, and the metallic lines disappear. Only in exceptional cases is it possible to make a gas luminous except by the electric discharge.

It would seem reasonable *a priori* to imagine that every molecule or atom has its own definite rate or rates of vibration, just as a tuning fork has one definite period and a piano wire several. In a solid or liquid constraints and collisions may produce forced vibrations covering a wide range of periods, but in gases or vapors, where molecular collisions must be comparatively few, there is a preponderance of free vibrations. If a molecule or atom has one free period of vibration, say that corresponding to the color of luminous sodium vapor, there will be but one image of the slit if the *D* lines were single—yellow in the case assumed. If there are a number of coexisting vibrations of different periods there will be an equal number of spectral lines.

504. Line Spectra are given by the metals and salts of the sodium and calcium groups in the Bunsen flame, and also by a number of other metals if spray from solutions of their salts, or ions caused by the electric spark, are passed into the flame. The spectrum of the electric spark or arc between electrodes composed of or coated with any metals or their salts contains many more lines than that of the flame. The number may be very large, ranging from a dozen or so in the case of the alkali metals to many thousands, in the cases of iron and uranium.

It seems that no two elements have any common lines, but the spectrum of a given element may show differences in the number and the appearance of the lines according to the nature of the source, whether flame, arc, or spark. (See Figs. 2, 3, Plate opposite p. 455.) All salts of the same metal give the same line spectrum, although in some cases they give band spectra as well, which may be characteristic of the salt (see next section). The lines of the non-metallic components do not appear with those of the metal except in rare cases. From the fact that a salt gives the spectrum of the metallic component alone, it would appear that line spectra are characteristic of the atomic state of the element and that metallic atoms are more easily excited to luminosity than non-metallic. Intense electric discharges through a non-metallic gas at ordinary pressures or in vacuum tubes give line spectra.

505. Band Spectra are usually composed of fine lines, as shown in Fig. 4 (the spectrum of part of the carbon arc, supposed to be due to cyanogen). The spectrum of the green cone in a Bunsen flame gives a very similar spectrum, due to carbon or its compounds. The salts of the calcium group of metals have flame spectra containing both lines and bands. All salts of calcium, for example, give the same flame spectrum under ordinary conditions,

but Mitscherlich (1862) found that if calcium chloride, for instance, is placed in a flame supplied with hydrochloric acid an entirely different band spectrum is produced, and still another if the flame is supplied with calcium bromide and an excess of hydrobromic acid. The inference is that in these cases the bands represent the characteristic spectra of the compounds, and that the spectrum observed under ordinary conditions is that of the oxide, due to reaction with atmospheric oxygen.

Nitrogen gives a band spectrum very similar to that of cyanogen if a feeble discharge passes through it, but an entirely different line spectrum if the discharge is very intense. (Figs. 5, 6.) Nitric oxide gives a characteristic band spectrum in the ultra violet very similar to that of nitrogen. All the compounds of mercury with chlorine, bromine, or iodine, give characteristic band spectra with feeble discharges, and the line spectrum of mercury with strong discharges. The same is true in a number of other cases. All these facts are consistent with the view that band spectra are characteristic of the molecular state of either elements or compounds. Intense discharges, by dissociating the molecule, will produce line spectra, characteristic of the atomic or, rather, "ionic" state. The salts of the alkali metals are so easily dissociated that they give only line spectra in the flame. It is possible, however, by other modes of excitation to produce band spectra of these elements (see § 526).

506. Structure of Lines. Not all spectral lines are the same in appearance, not even those due to the same element. There may be wide differences in intensity, width, and variability with changed conditions. It is evident that wide lines are not monochromatic, as they are practically small strips of a continuous spectrum. This is likewise true to a lesser extent of the narrowest lines. It has been shown by the use of instruments of great resolving power, such as the Fabry and Perot interferometer, that some apparently very narrow lines are highly complex, consisting of a number of narrow components, and no very bright line has been found that is due absolutely to monochromatic light.

At low pressures the spectrum of hydrogen due to the spark is composed of narrow lines, which become very broad at atmospheric pressure, with a background of continuous spectrum. This is probably due to more frequent collisions, which modify the free period, the conditions approaching those of an incandescent solid. An intense Leyden jar discharge through a vapor gives very broad lines, which may become quite narrow if self-inductance is put in the circuit. Increase of the pressure of the surrounding atmosphere often broadens the lines of a vapor which does not itself increase in density. Another remarkable effect produced by increased pressure in many instances is a very small shift of some of the lines, not necessarily all, toward the red. Increased temperature of the source generally broadens

the lines. This is probably in part at least due to the Doppler effect. The waves coming from the atoms approaching in the line of sight are shortened, those from receding atoms lengthened, by the telescoping or the stretching of the waves, just as the pitch of a locomotive whistle is apparently raised or lowered by the approach or receding of the train (§ 372).

507. Limits of the Spectrum. The very short ultra-violet waves are absorbed by all gases except hydrogen, and by most lenses and prisms. Working with fluorite lenses and prisms or a grating in a vacuum, Schumann and Lyman have reached a wave length of about .0001 mm. The ordinarily used unit of wave length is the Ångström unit, equal to one ten-millionth of a millimeter. This is sometimes called a tenth-meter. Another unit frequently used is the micron, $\mu = 0.001$ mm. Expressed in these units, some wave lengths are given below. Most substances are opaque to very long waves, and the longest waves mentioned were obtained by the method of selective reflection described in § 550, the wave length being then measured by a coarse grating.

	Ångström Units	μ
Shortest ultra-violet waves	1,000	0.1
Shortest visible waves (violet), about	3,800	0.38
Violet, about	4,000	0.4
Blue	4,500	0.45
Green	5,200	0.52
Yellow	5,700	0.57
Red	6,500	0.65
Longest visible waves (red)	7,500	0.75
Longest waves in solar spectrum, more than	53,000	5.3
Longest waves transmitted by fluorite	95,000	9.5
Longest waves by selective reflection from rock salt	500,000	50.0
By reflection from potassium chloride	612,000	61.2
Shortest electric waves	40,000,000 = 4 mm.	4000

ABSORPTION OF RADIANT ENERGY.

508. General Absorption. When radiation falls on matter a portion is reflected, another absorbed, and if the substance is transparent or very thin a third is transmitted. Black substances, such as lampblack and copper oxide, reflect and transmit very little, the absorption being almost complete. Most substances black to visible radiation are also black to the ultra-violet and infra-red waves, but there may be exceptions—for example, a

sheet of hard black rubber is opaque to visible radiation, but transparent to waves beyond the red (see § 311). Substances like that last mentioned, which absorb certain radiations and transmit others, are said to exercise selective absorption.

509. Selective Absorption is characteristic of most substances. Familiar examples are red glass, which transmits red and some infra-red, but no other visible colors; blue cobalt glass, which transmits blue and violet and a little red and green in narrow regions; green, which transmits almost all the colors, but a larger proportion of green; chlorophyll solution, potassium permanganate, the aniline colors, and solutions of the rare earths, didymium, etc. In most cases the absorption bands are wide and diffuse; in the case of the rare earths they are almost as narrow as spectral lines, so that the solutions appear almost colorless, no large amount of any one color being absorbed; the vapors of iodine, nitrogen peroxide, and some other substances have fluted absorption bands, grouped somewhat like the lines in the nitrogen bands. Many substances are very transparent within wide limits, beyond which they are completely opaque; for example, glass is opaque to waves shorter than 3,500 Ångström units, and longer than about 30,000 Ångström units. Quartz is transparent between the wave lengths 1,800 and 70,000, and for some longer waves; rock salt is transparent between 1,800 and 180,000, and fluorite, one of the most transparent substances, will transmit the shortest known ultra-violet waves ($\lambda = 1{,}000$) and up to $\lambda = 95{,}000$.

510. Kirchhoff's Law. If the fraction a of the radiation of a given wave length incident on a body is absorbed, a is said to be the coefficient of absorption of that color. The *emissivity* of a radiating body is the amount of energy radiated per second from each unit of surface. Kirchhoff showed by the theory of exchanges (§ 308) that the emissive and absorptive powers of all bodies at the same temperature for a given color are proportional when the radiation is a pure temperature effort.

511. Origin of the Fraunhofer Lines. Fraunhofer noticed that two bright yellow lines in the spectrum of a candle flame coincide with the D lines of the solar spectrum. For many years the origin of the Fraunhofer lines was disputed. It had been found that various colored vapors, such as nitrogen peroxide,

cause dark absorption bands in the spectrum somewhat similar to the solar lines, and it was generally believed that these also must be due to absorption; but whether this absorption took place in the sun's or the earth's atmosphere was unsettled. Finally Brewster (1860) gave good reasons for believing that some of these lines are due to absorption in the sun's atmosphere, and others to the earth's atmosphere. About the same time Kirchhoff, noting that there were coincidences between many of the Fraunhofer lines and emission lines, explained them as the result of absorption by vapors in the sun's atmosphere of waves which these vapors emit themselves. Before this, in 1849, Foucault had performed an experiment that had been overlooked, but which showed conclusively to what the lines are due. He noted that the yellow lines in the arc spectrum coincided with the *D* lines; that when the sun's light passed through the arc the *D* lines became darker; and, finally that the *D* lines could be reproduced by allowing the light from the white hot positive carbon to pass through the yellow flame. Foucault did not know the origin of the lines, but concluded that the medium which emitted and absorbed the *D* lines in the arc spectrum must exist in the sun. About the same time Stokes independently suggested that the coincidence of the yellow sodium lines with the *D* lines indicated that the sodium atoms must absorb waves of the same frequency as those emitted by them, the effect being similar to resonance phenomena in sound. This *reversal* of the sodium lines is easily secured by igniting a small piece of metallic sodium in a metal spoon before a slit illuminated with the electric arc, the light then passing through a prism and a lens which focuses it on a screen. If a large quantity of sodium vapor is present in an arc the phenomenon of *self-reversal* is shown in the spectrum. The bright lines are very broad and intense, with a narrow dark line in the middle of each, due to absorption by the cooler sodium vapor in the outer portion of the arc.

512. Luminescence. In all cases where radiation is purely a temperature effect Kirchhoff's law appears to hold. In many cases, such as those of fluorescence and phosphorescence (§§ 525, 526), in which the absorption of waves of certain lengths causes the emission of waves of a different length, this is not true; nor is it generally true of luminous gases and vapors, where the lumi-

nosity appears to be due to electrical or chemical causes. In no known case do gases or vapors have absorption lines corresponding to all the emission lines. The name *luminescence* has been applied to the various kinds of radiation not directly due to high temperature and not conforming to Kirchhoff's law.

513. Solar Spectrum. In 1888 Rowland published an enlarged series of photographs of the solar spectrum taken with a concave grating, the entire spectrum being about thirteen meters in length. The wave lengths of many thousands of the Fraunhofer lines were determined by him. A large number were found to coincide with the emission lines of known elements, so that it seems certain that about forty of these elements exist in the sun. The chromosphere, or gaseous solar atmosphere, the prominences or flames of incandescent hydrogen and other gases rising out of it, and the corona, or nebulous outer envelope, give bright line spectra, which may be seen during a total eclipse, when the brighter light from the photosphere does not mask them. The rare gas helium was known to exist in the sun before it was found on the earth, on account of the bright yellow line due to it observed in the spectrum of the prominences.

The ultra-violet region of the solar spectrum does not extend beyond a wave length of about 3,000 Angström units. Without doubt shorter waves are emitted, but they are absorbed by the earth's atmosphere, which is opaque to all very short waves. The atmosphere also exercises general and selective absorption in the visible region. Oxygen and water vapor give rise to the terrestrial lines and bands known as the Fraunhofer lines A, a and B, and there is more or less general absorption due to these and other constituents of the earth's atmosphere.

The infra-red region of the solar spectrum has been investigated by Langley with the bolometer, and found to extend beyond a wave length of 53,000 Ångström units. Broad absorption bands are found, some of which

Wave Lengths of Fraunhofer Lines.

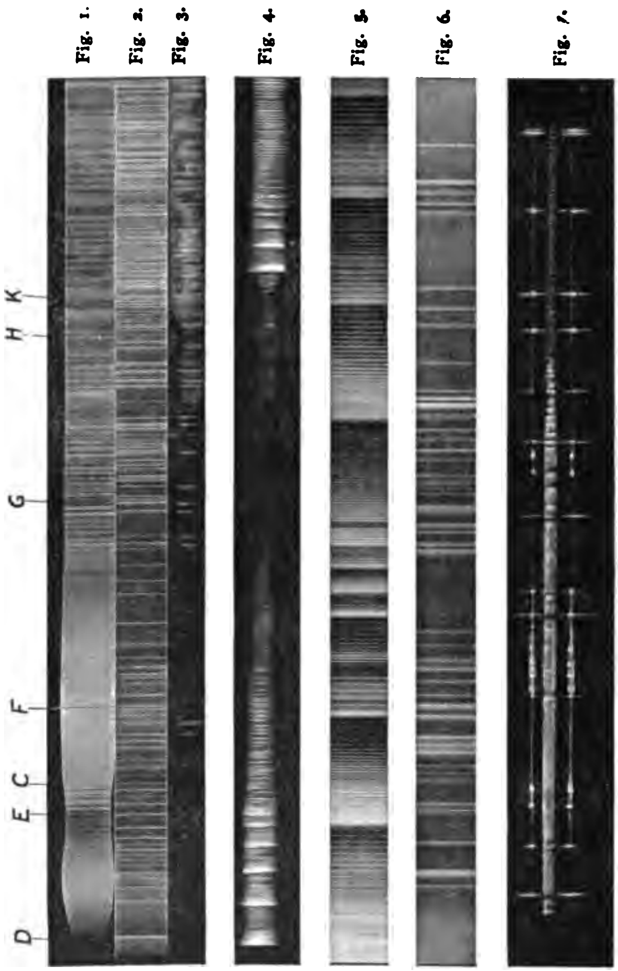
A	7594–7621	O	Red
B	6870	O	Red
D_1	5896.15	Na	Orange
D_2	5890.18	Na	Orange
E_2	5269.72	Fe	Green
F	4861.50	H	Blue
g	4226.89	Ca	Violet
H	3968.62	Ca	Violet
K	3933.81	Ca	Violet

coincide with those due to water vapor and carbon dioxide, besides many narrow lines and bands of unknown origin. A large proportion of the solar radiation, particularly in the neighborhood of the shorter waves, is absorbed by the earth's atmosphere, and this must greatly influence climatic conditions.

514. Spectra of Planets, Stars, Comets, and Nebulæ. The planets and the moon give spectra similar to that of the sun, as might be expected, but modified by general and selective absorption in the cases of the planets which have an atmosphere. Most stars have characteristic absorption spectra resembling that of the sun, which shows the universal distribution of many of the common elements. In addition there are frequently lines due to unknown elements. Some stars have bright as well as dark lines in their spectra. In such cases a considerable portion of the radiation must be due to the outer gaseous envelope or chromosphere, and these stars are probably in a condition more nearly resembling nebulæ than our sun, in which the denser photosphere emits a continuous spectrum brighter than that of the chromosphere, which absorbs more than it emits. Nebulæ give bright line spectra, some of the lines being due to hydrogen and helium, while others have not yet been identified. The spectrum of comets consists mostly of the characteristic hydrocarbon bands similar to those given by the green cone of the Bunsen flame. It seems evident in the cases of nebulæ and comets that the radiation is an example of luminescence, or luminosity due to other causes than high temperature, because these bodies appear to consist of masses of highly attenuated gases, or small bodies, and it is inconceivable that their temperature can permanently remain much higher than that of the surrounding space.

515. Applications of Doppler's Principle. If a star is approaching or receding from the earth with the velocity v the effect will be to shorten or lengthen each wave reaching the earth by an amount vT, if T is the period of vibration of the wave. If V is the velocity of light, $\lambda = VT$. The modified wave length is $\lambda_1 = (V \mp v)T$. The change in wave length is $\lambda_1 - \lambda = vT = (v/V)\lambda$. Each spectral line will be displaced by an amount expressed by the above relation, toward the violet if the star is approaching, toward the red if it is receding. By measuring such displacements on photographs of stellar spectra the velocities of stars in the line of sight may be determined with an error of less than one kilometer per second. Most of the stars which have been investigated have velocities with respect to the sun of between one and one hundred kilometers per second. It is found that a majority of the stars on one side of the heavens have a general relative motion toward the sun, those on the opposite side away from the sun. The inference is that the solar system is itself moving through the universe in the former direction. In many cases stellar lines become double, separate, and come together again in regular periods, while in other cases the lines remain single, but oscillate slowly about a mean position.

PLATE I. (Facing page 455.)



Explanation of Plate.

Fig. 1. Solar spectrum.

Fig. 2. Arc spectrum of iron. Many lines of solar spectrum coincide with these.

Fig. 3. Spark spectrum of iron. Note differences from above.

Fig. 4. Cyanogen bands, carbon arc spectrum.

Fig. 5. Band spectrum of nitrogen.

Fig. 6. Line spectrum of nitrogen.

Fig. 7. Spectrum χ Draconis, with comparison hydrogen spectrum above and below, showing Doppler effect (taken at Lick Observatory).

In the first case the light evidently comes from a double star. As the two components revolve about their common center of gravity, one will be approaching while the other is receding. This causes displacements in opposite directions, except when both stars are moving at right angles to the line of sight, when the lines coincide. In the second case the same condition must prevail, except that one component is dark. At intervals of about sixty-nine hours the variable star Algol has minima of brightness, probably due to its eclipse by a dark companion. The oscillations of its spectral lines have the same period and lead to the same conclusion. The orbital velocity of Algol determined from the maximum displacement of the lines is about forty-two kilometers per second. The period of the sun's rotation and that of the rotation of Saturn's rings has been determined from the relative displacements of the Fraunhofer lines observed in the spectra of the approaching and receding limbs. Sometimes there are dislocations of the lines due to the solar prominences, indicating that they are outbursts of incandescent gases projected with enormous velocities from the sun. The greater part of the work of some observatories, notably the Lick and Potsdam observatories, is based on the application of Doppler's principle to line-of-sight work and the discovery of "spectroscopic binaries," which sometimes cannot in any other way be shown to be double. (See Fig. 7, Plate I.)

EFFECTS DUE TO ABSORPTION.

516. Color of Natural Objects. The colors seen in the spectra produced by dispersion or by interference are pure. This is not the case with the colors of natural objects, which as a rule are due to selective absorption of certain colors of the incident light, the other colors being diffusely reflected in greater or less proportion. If a colored object, such as a red rose, is placed in different parts of a spectrum, it will appear a brilliant red in the red and almost black in other parts. This shows that the greater part of all colors except red is absorbed; not all, however, for it will be noticed that in every part of the spectrum there is some reflection of the incident color. Since the resultant of the combination of all colors is white, it may thus be proved that from all colored objects some white light is reflected, in addition to the characteristic color.

517. Body Color. In most cases it is observed that bodies having a certain color by reflected light have the same color by transmitted light. This suggests that the color diffusely reflected is due to components of the incident white light which have pene-

trated more or less into the medium before being scattered, the other colors being lost by absorption. The white light reflected is probably due both to reflection at the surface and to the recombination of the various colors which escape complete absorption. As a crude illustration of body color, if light falls on a piece of red glass a white image of the source will be reflected from the front surface and a red image from the rear surface.

Colors are said to be more or less *saturated* according to the proportion of white light with which they are diluted. The pure spectral colors are said to be completely saturated. The proportion of white light scattered is increased by any process which increases the reflecting surface. For example, crystals of copper sulphate will appear lighter and lighter as they are crushed into smaller fragments, and become almost white when reduced to a fine powder. The white light reflected from the numerous surfaces then completely masks the small portion which is selectively transmitted. Similarly, transparent substances such as glass are white when in powdered form.

518. Dichromatism. Some substances when examined by light transmitted through thick layers appear to be of a different color from that observed by reflection or by transmission through a thin layer. A thin layer of chlorophyll is green by transmitted light, while a thick layer is red. This is explained by the fact that the coefficient of absorption, or the fraction of the incident light absorbed by a layer of unit thickness, is different for the two colors. While the incident green light is more intense than the red, and remains so after transmission through a thin layer, it is more rapidly cut down by absorption, so that after passing through a thick layer the red predominates. This effect is called dichromatism.

519. Surface Color. Some substances appear of different colors by reflected and by transmitted light. Such is the case with thin films of metal and of the solid aniline colors. Gold is always yellow by reflected light, but a sheet of gold leaf thin enough to permit transmission appears green by the transmitted light. The light reflected from these substances is complementary to that transmitted. In such cases selective action seems to take place at the surface, some colors being directly reflected, others being

absorbed by a thick layer, or transmitted through a thin film. Bodies exhibiting surface color retain that color when finely powdered.

520. Colors of Sky and Clouds. Since light can reach the eye only directly from the source or by reflection from material objects, it is evident that, since the sky is not perfectly black, it must contain matter in suspension. Some have supposed that air itself may have a characteristic color, as is shown by great thicknesses of glass or of water, but the experiments of Tyndall and others indicate that the blue color of the sky is due to selective scattering by small suspended particles of dust, water, etc. It is to be expected that such small particles should reflect a larger proportion of short waves than of long ones. The term scattering is used, because it seems evident that this is not a case of ordinary reflection like that from a mirror of finite size. There is an analogy in the case of sound waves; long waves pass around obstacles without deviation from their general direction, while shorter waves may be reflected. Since the shorter waves of light are scattered, the transmitted light will consist mostly of the longer waves. This accounts for the brilliant reds, oranges, and greens often observed in the western sky at sunset. The light transmitted almost tangentially through the atmosphere has been deprived of the shorter waves, which cause a blue sky for those more immediately under the sun. These effects are intensified by the presence of a large number of dust particles in the lower levels of the atmosphere. After the great eruption of the volcano Krakatoa in 1883 fine volcanic dust pervaded the atmosphere of the whole earth and the sunsets were especially brilliant. For the same reason lights look red when seen through smoke or fog, or through water made slightly turbid by the addition of a small quantity of milk or shellac solution. This effect is beautifully illustrated by passing a beam of light through a jet of steam issuing from a small nozzle into a stream of air previously dried by forcing it through sulphuric acid. The size of the water drops is controlled by changing the vapor pressure in the atmosphere in which the drops are formed, lower vapor pressure promoting evaporation and thus reducing the size of the drops. The colors

seen by transmitted and by scattered light are complementary, the shorter waves being scattered and the longer ones transmitted.

521. Color Sensation. The perception of a given color by the eye does not necessarily prove that the stimulus is of the corresponding wave length. It may be the resultant effect of several different colors. For example, if the light from the red of a spectrum and from a region intermediate between the blue and the green be superimposed the resultant sensation is white, which the eye cannot distinguish from the white due to a mixture of all the colors. A similar effect is produced by the combination of violet and yellow-green. Two colors which together give the sensation of white are said to be complementary. It is found, further, that spectral red and green combined excite a sensation of yellow, while green and violet produce blue. All possible colors may be produced by combining red, green, and violet. According to the theory of Thomas Young, adopted and extended by Helmholtz and by Maxwell, these are to be regarded as the three primary color sensations. The cones in the retina are supposed to respond or "reso nate" most actively to frequencies of vibration corresponding to these

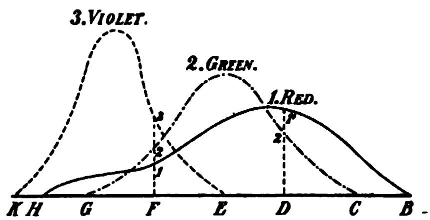


FIG. 329.

colors, and all color sensations depend on the proportions of the incident energy belonging to these frequencies. The phenomena of color sensation may be explained by assuming that in the normal eye there are three sets of nerves, one stimulated most actively by red light, another by green, and a third by violet, but each responding also more or less to waves of other frequencies as well. To indicate these effects Koenig constructed three curves (Fig. 329) on an axis representing the length of the visible spectrum from the Fraunhofer line *K* in the violet to *B* in the red. The ordinates at each point represent the degree of excitation of the three sets of nerves respectively by light of the frequency corresponding to that point in the spectrum. The maximum sensibility of the "red" nerves is in the orange-red, that of the "green" nerves in the green, and that of the "violet" nerves in the blue-violet. The first set of nerves is also excited more or less by all colors between *H* and *B*; the green by all colors between *G* and *C*, and the violet by all colors between *K* and *E*. The color of sodium light

(D) is caused by the superposition of two sensations, red and green, proportional respectively to the ordinates $D1$ and $D2$. The color of the blue line of hydrogen (F) is due to a combination of red, green, and violet sensations proportional to the ordinates $F1$, $F2$, and $F3$. In the case of color blind persons one or more sets of nerves is missing—usually the red. To such persons, for example, sodium light would appear green. Two colors, such as red and blue-green, are complementary, when, acting jointly, they excite all three sets of nerves in the proper proportion to produce the sensation of white—or in the same proportions that they are excited by ordinary white light.

522. Pigment Colors. The effect of mixing pigments is quite different from that of mixing spectral colors. For example, blue paint absorbs nearly all the incident light except the blue and some green; yellow paint absorbs nearly all except yellow and some green. If, therefore, white light is incident on a mixture of the two pigments green is the only color which escapes absorption by one or the other, therefore a mixture of blue and yellow paints produces a green paint. In such cases the apparent color may vary with the kind of illumination. Blue pigments usually appear green by candle light, because there is a very small proportion of blue in the incident light, and so green predominates in the scattered light.

523. Tyndall pointed out a transformation which he called **calorescence**, in which longer waves are changed to shorter. Solar or arc radiation is passed through a cell filled with iodine dissolved in carbon bisulphide, which is opaque to visible radiation but not to the longer waves. If the transmitted radiation is focused by a lens it will raise a piece of metal to red heat, thus giving rise to visible radiation. In all cases where radiation is dependent on temperature alone Kirchhoff's law applies, and there is a definite ratio between the emission and absorption of a body. The above case is no exception; although the incandescence is produced by long invisible waves, the heated body can never be brought to the temperature of the source, the sun or arc, and the general shape of the intensity curve is like that of the source. The luminous energy forms such a small part of the whole that its absorption by the iodine scarcely influences the result.

524. Chemical and Molecular Effects. Light may cause chemical combination, as when it acts on a mixture of hydrogen and chlorine, or dissociation, as when it acts on the silver salts in a photographic plate. By its action on the chlorophyll of plants, light decomposes the carbon dioxide absorbed from the atmosphere releas-

ing the oxygen and causing the carbon to be assimilated. It may cause molecular transformations, as when it alters amorphous to crystalline selenium, or changes the electric resistance of the latter form. It also changes white phosphorus to red. These effects are not due to the heating effect of the absorbed radiation, because an equivalent rise of temperature will not cause them, but they seem rather to depend on the vibratory character of the light waves. The molecules or atoms of the affected substances appear to resonate to certain wave frequencies, and the observed changes result from the inter-molecular agitation. As a rule the shorter waves are the most effective in producing such results. From this fact and from that of the greatest heating effect being due to the long waves, arose the old arbitrary distinction between visible, actinic, and heat radiation.

Another effect due to light, especially to the ultra-violet waves, is that it will cause the discharge of electricity from certain metals when they are polished, insulated, and charged with negative electricity. In this case electrons, or small negatively charged corpuscles, appear to be shaken out of the metal by the rapid light vibrations (§ 747).

525. Fluorescence. There are substances which when stimulated by the absorption of waves of certain lengths will emit waves of different lengths. For example, a piece of paper moistened with sulphate of quinine solution and held in the ultra-violet portion of the solar spectrum will emit a brilliant opalescent blue light. This effect was discovered by John Herschel in 1845, and also studied by Brewster in 1846. To this phenomenon Stokes gave the name of fluorescence, because it was observed in fluorspar. He explained it as the result of the absorption of incident waves which by a modified resonance action caused a reëmission of longer waves. Similar effects are observed in coal oil, fluorescein, eosin, uranin, and other organic compounds; in uranium glass, which emits a yellowish green light; in esculin, which emits blue light, and in chlorophyll, which emits red light; and also in a much smaller degree in iodine, wood, paper, and many other substances.

526. Phosphorescence. There is a large class of substances, of which calcium, strontium, and barium sulphides are familiar examples, which after exposure to light show effects which are

similar to fluorescence, but which continue visible long after the exciting radiation ceases to act. This is called phosphorescence, a name derived from the similar appearance of phosphorus while slowly oxidizing. The only definite distinction between fluorescence and phosphorescence is that the latter persists for a longer time. As found by Dewar, many substances which phosphoresce very feebly at ordinary temperatures may be made to glow brilliantly at the temperature of liquid air. As examples, gelatin, horn, egg shells, and paper may be mentioned.

Similar effects can be excited in many substances by other causes than by light—by chemical action, as in the case of phosphorus exposed to oxygen; by electrical action, as in Crookes' tubes; by mechanical action, as seen in sugar broken in the dark; by elevation of temperature, as in the case of the diamond; and by Röntgen or Becquerel rays (§ 736).

Some metallic vapors, such as those of the sodium and calcium group, fluoresce brilliantly under the action of light or cathode rays. The light shows the characteristic spectral lines and bands of the metal. Certain organic vapors, such as anthracene, fluoresce when light falls on them. Nitrogen, oxygen, and some other gases will under certain conditions phosphoresce brightly for several seconds after an electric discharge has passed through them in a vacuum tube (§ 731).

DOUBLE REFRACTION AND POLARIZATION.

527. Double Refraction. Some crystals, such as those of rock salt and fluorite, resemble isotropic solids, such as glass, in the respect that their physical properties are alike in all directions. In general, however, this is not the case; such properties as elasticity and heat conduction, as well as optical properties, differ in different directions in the crystal. In such crystals as quartz and calcite there is an axis of symmetry, the crystallographic axis, and the physical properties are the same in all directions in any equatorial plane, but different from those in the direction of the axis. Iceland spar, or calcite, is a rhombohedral crystal, each face being a parallelogram with two acute angles of 78° 5′ and two obtuse angles of 101° 55′. Two solid angles of the crystal are formed by the junction of the obtuse angles of three faces. A line

symmetrically situated with respect to these solid angles is a crystallographic axis. In 1669 Bartholinus, discovered that an object seen through Iceland spar appears double, unless viewed in the direction of the axis. No such effect is observed in the case of isometric crystals. This phenomenon is called *double refraction*. When the waves travel in the crystal in the direction of the crystallographic axis there is no double refraction; hence any line in the crystal parallel to the axis is called an *optic axis*.

These experiments were repeated and extended by Huyghens in 1690, and partly explained by him on the basis of the wave theory. If a ray *r* of ordinary light is incident normally on any face of a doubly-refracting crystal one ray *o* is transmitted without devia-

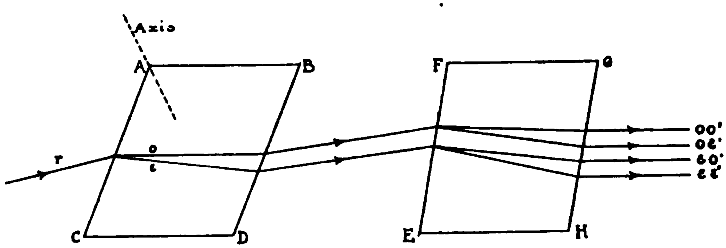


FIG. 330.

tion; and if the incidence is oblique (Fig. 330) this ray is deviated, with an index of refraction which is independent of the angle of incidence. The other ray *e* is deviated in all cases, unless it travels along an optic axis, and the index of refraction varies with the angle of incidence. The first is called the *ordinary*, the second the *extraordinary* ray. If the crystal be rotated, keeping the angle of incidence constant, the ordinary image will remain at rest, while the extraordinary image rotates about it in such a way that the line joining the two images lies in a *principal section*, a plane including the normal to the surface and an optic axis. If the ordinary and extraordinary rays *o* and *e* pass through a second crystal each ray generally divides in two, the rays *oo'* and *oe'* and the rays *eo'* and *ee'* (Fig. 330), the line joining each pair lying in a principal section of the second crystal. This gives rise to four images of the source, which are of equal intensity when the prin-

cipal sections of the two crystals are at an angle of 45° with each other. If this angle be changed one pair of images will increase in intensity and the other diminish. When the principal sections are parallel only the rays *oo′* and *ee′* emerge; when they are at right angles, only the rays *oe′* and *eo′*. From such experiments Huyghens recognized the fact that light which has passed through Iceland spar, quartz, and other doubly-refracting crystals does not possess properties which are alike in all azimuths around the direction of propagation. Newton, in order to explain this, supposed the light corpuscles to be endowed with polarity of some sort—hence the name polarized light.

528. The Wave Surfaces. From the experiments described above it may be seen that a wave of ordinary light on entering a doubly-refracting crystal is divided into two waves, one of which, *o*, has the same velocity in all directions in the crystal. The other wave *e* has a velocity which varies in different directions, and is the same as that of the ordinary wave only when both travel in the direction of the optic axis. Huyghens showed that these facts are consistent with the existence of a double wave surface in the crystal, a sphere and an ellipsoid of revolution, which are tangent to each other at the two points where they intersect an optic axis. In one class of crystals, like Iceland spar, the sphere is inside the ellipsoid, and the ordinary wave is the more refracted (Fig. 331). In another class, represented by potassium sulphate or quartz, the sphere encloses the ellipsoid, and the ordinary ray is less refracted (Fig. 332). The first are called negative and the second positive crystals. Wollaston, Stokes, Glazebrook, and others have proved the correctness of this construction by experiment. In the case of quartz the two wave surfaces do not touch where they intersect the optic axis, and there is double refraction of another kind in the direction of the axis (§ 545).

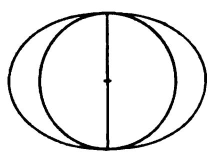


FIG. 331.

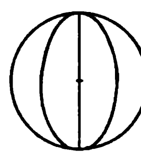


FIG. 332.

In crystals in which the physical properties are different along

three axes at right angles to each other, such as sugar and topaz, which likewise show double refraction, there are two axes of no double refraction; hence such crystals are said to be *biaxial*, as contrasted with the class described above, which are said to be *uniaxial*. As shown by Fresnel, both rays in biaxial crystals are extraordinary, that is to say, do not conform to the ordinary laws of refraction.

529. Double Refraction by Tourmaline. Tourmaline is a semi-transparent hexagonal crystal. If light falls on a crystal, part is transmitted. If this falls on a second plate with its axis parallel to that of the first (Fig. 333), some of the light gets through; but

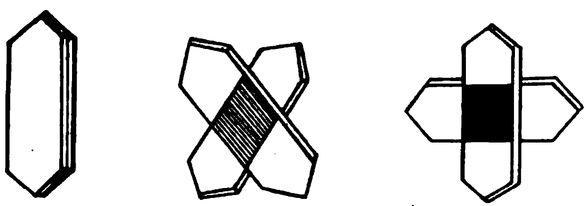


FIG. 333.

if the second crystal is rotated about the line joining the two, less light gets through, and when its axis is at right angles to that of the first none is transmitted. Evidently the waves have had their mode of vibration so changed by passage through the first plate that they cannot pass through the second unless the principal sections of the two are parallel. If the light first passes through Iceland spar it is found that the extraordinary ray alone will pass through tourmaline if the principal sections of the two crystals are parallel, the ordinary ray alone if they are at right angles. It follows that light is doubly refracted by tourmaline, but that the ordinary ray is totally absorbed. As a remarkable example of Kirchhoff's law (§ 510), it may be mentioned that if tourmaline is raised to a high temperature it emits polarized radiation. If this falls on a second crystal parallel to the first it is absorbed, showing that it corresponds to the ordinary ray. The mode of vibration which is absorbed corresponds to that which is emitted.

530. Polarization by Reflection. About 1808 Malus discovered

that light reflected from glass at a definite angle acquires properties similar to that of light transmitted through tourmaline or Iceland spar. When light is polarized by reflection from a mirror A (Fig. 334a) a large fraction is reflected from another mirror B if the two planes of incidence coincide. If the planes of incidence are at right angles (Fig. 334b) very little is reflected. If the light reflected from a glass plate is examined through a crystal of Iceland spar, the ordinary ray alone is transmitted when the plane of reflection coincides with a principal section, the extraordinary ray alone when the two are at right angles. In intermediate positions portions of both rays are transmitted. Similarly light reflected from glass is not transmitted through tourmaline if the plane of reflection is parallel to the optical axis of the crystal.

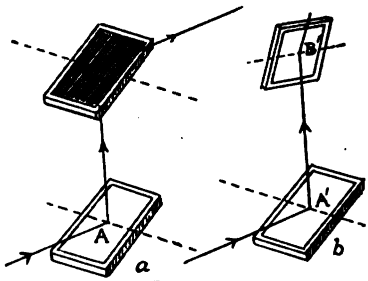


FIG. 334.

531. Direction of Vibration. Fresnel explained these phenomena as a result of transverse vibrations. In ordinary white light successive trains of waves reaching a given point of space vibrate in different planes at random, so that, although the waves in each train are vibrating transversely in a definite plane, and are, therefore, polarized, this direction changes so rapidly that the eye cannot take account of it and no polarization effects are observed. In passing through a doubly-refracting crystal vibrations in different directions travel with different velocities on account of the difference of the physical properties of the crystal in these directions, hence double refraction results. The displacement in each wave is in general resolved into two components, unless the light is travelling parallel to the axis. In that case it is unmodified, as velocity of propagation is independent of the azimuth. In the ordinary ray, which travels in all directions with the same velocity, the vibrations must be at right angles to the optic axis. So long as this is the case the displacements will take place under the same conditions in every azimuth and the velocity will be unchanged. In the extraordinary ray the vibrations

must be in a principal section. This accounts for the fact that the ordinary and the extraordinary image are always in a line parallel to a principal section. When light strikes a reflecting surface there is a partial resolution into components respectively in and at right angles to the plane of incidence. The vibrations parallel to the surface are most freely reflected, while the others strike down into the surface and are transmitted or absorbed (Fig. 335). If polarized light is incident at the angle of maximum polarization on a piece of glass a large proportion will be reflected when its vibrations are parallel to the surface; if the vibrations are in the plane of incidence it will be refracted.

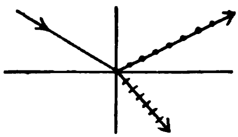


Fig. 335.

It is clear that no interference effects can be produced between two vibrations in planes at right angles to each other. This fact enabled Wiener to determine the direction of vibration in light polarized by reflection. A beam polarized by the mirror *M* fell at angle of 45° on a thin transparent photographic film above a reflecting surface. He found that stationary waves (§ 473) were produced when the plane of incidence on the film coincided with the plane of reflection from the mirror (Fig. 336 *A*), but this was not the case if the two planes were at right angles to each other (Fig. 336 *B*). From the figure it appears that in the first case the vibrations must have been parallel to the film, and therefore to the mirror, in order that the incident and reflected rays should be in a condition to interfere at *P*, while in the second case the vibrations must have been in the plane of incidence on the film, and therefore parallel to the mirror, in order that the vibrations should meet at *P* at right angles to each other. This demonstrates that the vibrations in light polarized by reflection are parallel to the mirror.

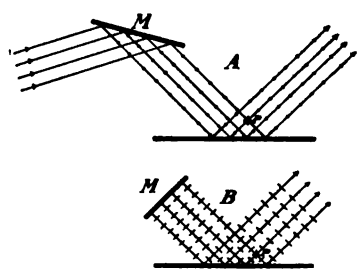


Fig. 336.

532. Plane of Polarization. Before the direction of vibration in polarized light was known it became customary to speak of the "plane of polarization" of a polarized beam, rather than of

the direction of vibration, and this plane was so defined that it coincides with the plane of incidence when the light is polarized by reflection. It follows that the vibrations in a polarized beam are at right angles to the plane of polarization.

533. Brewster's Law. The light reflected from a surface is not in general completely polarized, that is, all its vibrations are not strictly in one plane. It is found, however, that for each reflecting substance there is a certain angle of incidence for which the polarization is a maximum. This is called the *polarizing angle.* It was found by Fresnel that complete polarization is given only by substances having an index of refraction equal to about 1.46. Brewster found that the polarizing angle is such that the reflected and the refracted rays are at right angles to each other. Since $n = (\sin i)/(\sin r)$ and since, when $i = p$, the polarizing angle, $p + r = 90°$,

$$n = (\sin p)/(\cos p) = \tan p$$

From this relation the polarizing angle p may be determined. This is known as Brewster's law.

534. Pile of Plates. Since only a small fraction of the incident light is reflected from a transparent substance, even when the reflected light is completely polarized, the refracted light will be only partially polarized; that is to say, along with that vibrating in the plane of incidence a considerable proportion of that vibrating at right angles to this plane will be transmitted. If the light is subject to a second reflection the proportion of polarized light is increased. After passing through eight or ten plates the transmitted light is almost completely polarized. If a pile P of thin glass plates is built up as shown in Fig. 337, the beam R, the result of successive reflections, and the beam T, which is transmitted, are completely polarized in planes at right angles to each other. This is one of the simplest methods of securing polarized light.

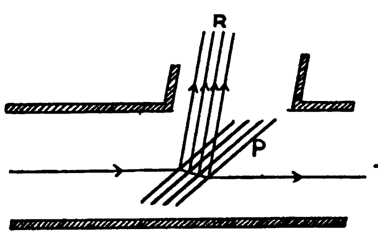


FIG. 337.

535. Wave Front Construction. If C is a radiant point in a crystal of Iceland spar (Fig. 338) and if AA' is the optic axis passing through that point, two waves will diverge from C, one spherical and the other spheroidal. These waves will have the same velocity along AA', but in other directions the extraordinary wave will travel faster than the ordinary. The vibrations in each wave will be in the wave surface. The vibrations in the ordinary ray will everywhere be at right angles both to the optic axis and to the direction of propagation.

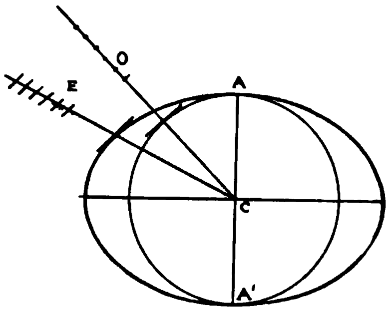


FIG. 338.

In the extraordinary wave the vibrations are in general oblique both to the optic axis and to the direction of propagation of the disturbance. In this case we have an exception to the general rule that the wave normal indicates the direction of propagation.

By the application of Huyghens' principle the wave fronts in double re-

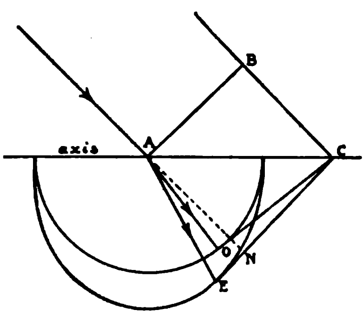


FIG. 339.

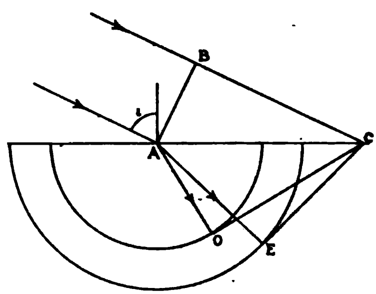


FIG. 340.

fraction may easily be determined. Consider a plane wave AB incident on a crystal so cut that the optic axis is parallel to the surface and to the plane of incidence (Fig. 329). The two disturbances in the crystal will travel to O and E respectively while the wave travels from B to C in air. The tangent planes CO and CE are the two wave fronts. The disturbance at E is due to A, a point not on the normal to the extraordinary wave front passing through E. The wave velocity, or the velocity of the wave front, is proportional to the normal distance AN; the ray velocity, or actual velocity of the disturbance, is proportional to AE.

When the axis is parallel to the surface, but at right angles to the plane of incidence, the wave front is found as shown in Fig. 340. In this case, the extraordinary wave also has a circular section. Only in this plane of incidence is the ratio $(\sin i)/(\sin r)$ constant for the extraordinary ray, and this ratio is called n_e, the extraordinary index of refraction. The value of the ratio V/V_e differs with the direction in every other plane of incidence, and hence cannot properly be called the index of refraction.

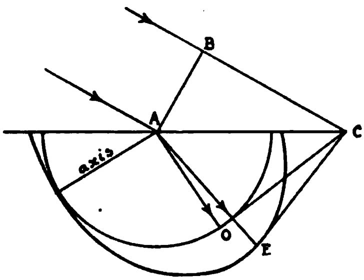


FIG. 341.

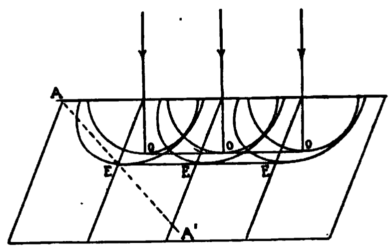


FIG. 342.

The general case, where the axis is at any angle with the refracting surface and the plane of incidence, is shown in Fig. 341.

When the wave is incident normally on any face of a crystal of Iceland spar the wave fronts are as shown in Fig. 342, which explains why the extraordinary ray is deviated.

536. Uniaxial Prisms. When light is incident on a doubly-refracting prism with its axis parallel to the base, there will be no separation if the light is transmitted at minimum deviation. If, however, the axis is parallel to the refracting edge (Fig. 343), the ordinary and the extraordinary rays will be separated, and the angular divergence will persist after emergence. Two spectra will be formed, with light polarized in opposite planes. The ordinary spectrum will be less deviated than the extraordinary by a quartz prism and more by a calcite prism. When the optic axis is parallel to the refracting edge of the prism the two indices of refraction may be determined from the relation

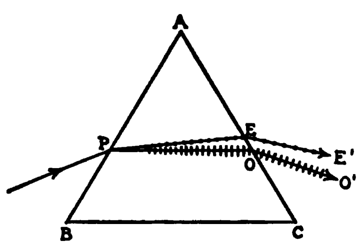


FIG. 343.

$$n_o = \frac{\sin \frac{A + D_o}{2}}{\sin \frac{A}{2}} \quad \text{and} \quad n_e = \frac{\sin \frac{A + D_e}{2}}{\sin \frac{A}{2}}$$

Some values of the index of refraction for sodium light are given below:

	n_o	n_e
Positive Crystals:		
Quartz	1.5442	1.5533
Ice	1.3091	1.3104
Negative Crystals:		
Calcite (Iceland spar)	1.6584	1.4864
Beryll	1.5740	1.5674
Sodium nitrate	1.5874	1.5361

The difference between n_o and n_e is greater in the case of Iceland spar than in any other ordinary crystal.

537. Polarizing Prisms. The two polarized rays produced by a doubly-refracting crystal are not sufficiently separated to be conveniently used when a single beam is desired. The separation may be increased by using an ordinary trihedral prism, but this introduces dispersion, so that other devices must be employed. The most common is the rhombohedral prism invented by Nicol, of Edinburgh, in 1828. In the principal section of a crystal of calcite (Fig. 344) the angles at *B* and *D* are 71°. The two end faces *AB* and *CD* are cut down to *A'B* and *C'D*, so that these angles are reduced to 68°. The crystal is then sliced along *A'C'* in a plane perpendicular to the ends and to the principal section. The two surfaces are polished and cemented together with Canada balsam, which has an index of refraction less than that of the calcite for the ordinary and greater for the extraordinary ray. If a ray of light *r* is incident in a direction parallel to the edge *AD* the ordinary ray will be totally reflected from the Canada

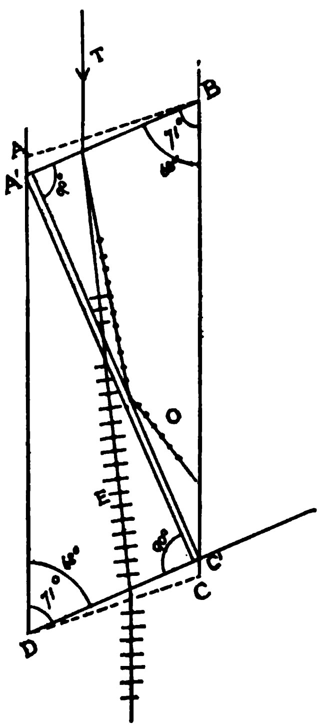


FIG. 344.

balsam, while the greater portion of the extraordinary ray will be transmitted. The reduction of the angles at A and D is for the purpose of securing the proper angle of incidence on the balsam to produce this effect.

The Foucault prism resembles that of Nicol, but the total reflection is from an air film. This allows the prism to be made shorter, but there is a greater loss of light by reflection and a smaller field of view.

538. The Polariscope is an instrument for the study of the optical properties of substances with respect to polarized light. It consists of two Nicol prisms or piles of plates, one called the *polarizer,* to produce the polarized light, the other, the *analyzer,* which may be set with its principal section at any desired angle with that of the polarizer, to test the incident light with respect to the nature and direction of its polarization. If any doubly-refracting substance is placed between the two its effects on the polarized light transmitted through it may be studied by the analyzer.

539. Resolution and Composition of Vibrations. If the polarizer and analyzer are set with their principal sections parallel, light which has traversed the first will pass through the second without sensible loss. If their principal sections are at right angles to each other, or "set for extinction," no light will be transmitted through the analyzer. If the angle between the principal sections is α (Fig. 345), and if a is the amplitude of the light transmitted by the first Nicol, the amplitude of that transmitted through the second is $a \cos \alpha$, and its intensity is proportional to $a^2 \cos^2 \alpha$. The intensity of the totally reflected ordinary ray is $a^2 \sin^2 \alpha$. The sum of the two intensities is $a^2(\cos^2 \alpha + \sin^2 \alpha) = a^2$, which is equal to the intensity of the light incident on the analyzer. This simple law of resolution of vibrations into components by double refraction, giving determinate control of the intensity through a wide range, is made use of in several forms of photometer.

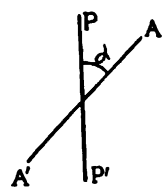


FIG. 345.

If the two Nicols are replaced by two crystals of calcite with their principal sections at an angle of α with each other, as in Huyghens' experiment (§ 527), an ordinary ray o and an extra-

ordinary ray e of the same amplitude a are produced by the resolution of the vibrations along two directions in the first crystal. At incidence on the second crystal, the ordinary ray will be resolved into the components oo' and oe' of amplitudes $a \cos \alpha$ and $a \sin \alpha$ and the extraordinary ray into the components eo' and ee', of amplitudes $a \sin \alpha$ and $a \cos \alpha$. There will be, therefore, in general four rays, as found by Huyghens, which will be of equal intensity when $\alpha = 45°$. When the principal planes are at right angles, the incident ordinary ray goes through the second crystal as an extraordinary ray and the extraordinary as an ordinary ray, and there are only two images.

If the second crystal is replaced by a Nicol prism, with its principal section parallel to that of the first crystal, only the components oe' and ee' emerge, their vibrations being in the same plane, that of the principal section of the analyzer. If the two rays are superimposed on emergence, the intensity will depend not only on the amplitudes of the two components, but on the phase differences which have been introduced owing to the difference in velocity in the crystal of the two rays from which these components are derived; in other words, there may be interference provided the light falling on the first crystal is plane polarized (see next section).

540. Interference of Parallel Polarized Light. If parallel plane polarized white light passes through a crystal of uniform thickness t and then through an analyzer, uniform colored effects are produced over the entire field, since some colors are reënforced and some weakened by interference. There is no real loss or gain for any color, for as shown in § 539, whatever energy is lost in the extraordinary ray is gained by the ordinary, and conversely, so that the ordinary light which is internally reflected in the prism is complementary to that passing through. When the principal sections of the crystal and the analyzer are either parallel or at right angles to each other no modification of the light is produced, the ordinary or the extraordinary ray alone getting through, so that there can be no interference. In all other positions of the analyzer there are varying proportions of white and colored light transmitted, the color effects being most pronounced when the principal sections are at an angle of 45° with each other.

The original beam of light falling on the crystal must be plane polarized. If ordinary light is used the succession of waves vibrating in different planes when resolved in the crystal will give rise to all possible distributions of amplitudes, so that all colors will be equally affected and the resultant effect will be white light.

541. Interference of Convergent or Divergent Light. If a divergent or convergent pencil of polarized light falls on a doubly-refracting crystal, different portions of the pencil will traverse the crystal at different angles, and therefore with different optical paths, hence the interference effects will not be uniform over the whole field. In general the effects are quite complex and cannot be discussed here, but the simple case of a uniaxial crystal cut perpendicularly to the optic axis may be considered as an illustration. Consider such a pencil diverging from S and falling normally on the face $ABCD$ of a doubly refracting crystal with its axis parallel to AA' (Fig. 346). The vibrations of the incident

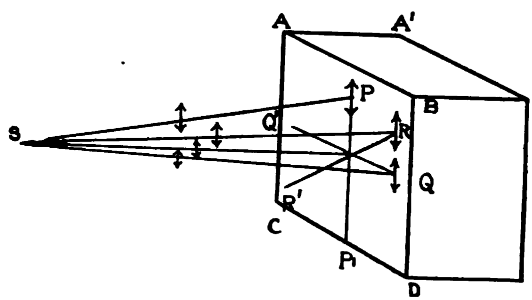


FIG. 346.

light may be supposed to be in a vertical plane, as indicated by the arrows. At P and Q the incident vibrations are respectively parallel and perpendicular to the principal sections PP' and QQ' and travel through without change. If an analyzer is placed beyond the crystal and set for transmission or extinction of light transmitted by the polarizer, there will be a light cross or a dark cross on a screen beyond it corresponding to the crossed lines PP' and QQ'. The light incident at such a point as R, however, will

FIG. 347.

be vibrating at an angle with the principal section RR', and will be resolved into two components. A relative difference of phase between them will exist at emergence, and interference effects will take place when they are re-resolved into the same plane by the analyzer. The same difference of path will exist for all rays incident at the same angle, that is, at all points equidistant from the normal from S to $ABCD$, hence colored rings similar to Newton's rings in appearance will be projected on a screen beyond the analyzer. The "rings and brushes" due to a calcite plate are shown in Fig. 347. The brushes are dark, showing that the Nicols are crossed.

The interference effects due to crystals cut in other ways or to biaxial crystals are analogous to those described above, but more complex.

542. Double Refraction due to Strain. If a plate of glass or other isotropic substance is placed between a polarizer and an analyzer set for extinction no effect is produced. If the substance is then compressed or stretched some light will pass and interference effects similar to those described above will be produced. This shows that an isotropic substance becomes doubly refracting when subjected to unsymmetrical strain. This method offers a very delicate test of deviations from isotropy. Some liquids show the same characteristics in cases where the viscosity is so great or the stress so suddenly applied that a uniform hydrostatic pressure has not had time to become established throughout the substance. Imperfectly annealed glass exhibits double refraction. As shown by Tyndall, a bar of glass set in longitudinal vibration restores the light through the crossed nicols, and a rotating mirror shows that the effect is set up periodically as the compression waves pass across the field.

Kerr found that a block of glass in a strong electrostatic field becomes doubly refracting like a uniaxial crystal with its axis parallel to the field (§ 611).

543. Circular and Elliptical Polarization. Consider the state of the light originally plane polarized as it emerges from a doubly-refracting crystal before it reaches the analyzer. The ordinary and extraordinary rays start from the first surface in the same phase, but, as their velocities are different, one set of waves will fall behind the other. At different points within the crystal the ether will be subject to two disturbances at right angles to each other and with phase differences depending upon the thickness of the medium traversed. The optical difference of path d at a distance t from the first surface is $[(V/V_e) - (V/V_o)]t$. At points where this difference is $n\lambda$ the light is plane polarized in a direction intermediate between the planes of vibration of the two components, the slope depending on their relative amplitudes, and being 45° if these are equal. If the difference of path is $(2n + 1)/2 \cdot \lambda$ the light will likewise be plane polarized, but with a reversed direction of slope. If the difference of path is any odd multiple of a quarter of a wave-length the disturbance will be elliptical, or circular if the amplitudes are equal. For intermediate differences of path the disturbance will be elliptical, the axes of the ellipse being oblique with respect to the axis of the crystal. The successive stages at different distances from the first surface are shown in Fig. 348. On emergence from the crystal the ether disturbance will preserve the final form, and will be plane, elliptically, or circularly polarized according to the thickness of the crystal. If the waves are circularly polarized the disturbance travels through space like a point on a rotating screw. The polarization is said to be right-handed if the rotation is clockwise looking in the direction of propagation, or if the displacement resembles that of a right-handed screw, left-handed if the displacement is like that of a left-handed screw.

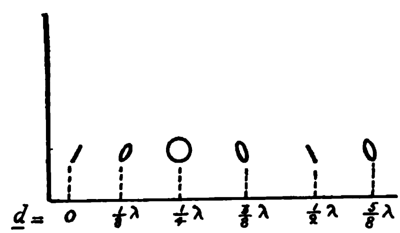


FIG. 348.

When light is totally reflected there is a phase difference between the vibrations respectively in and at right angles to the plane of incidence, so that this light is elliptically or circularly polarized. In ordinary reflection there is a slight elliptical polarization, which becomes very marked in the case of metallic reflection.

544. Production and Detection of Elliptically Polarized Light. Cir-

cularly or elliptically polarized light cannot be detected by the unaided eye. If viewed through a Nicol prism no change in the intensity of circularly polarized light accompanies rotation of the prism, as a component of unchanging magnitude is transmitted. If the light is elliptically polarized there will be variations of intensity as the prism is rotated, the intensity being greatest when the principal section of the prism is parallel to the major axis of the ellipse (component amplitude of greatest magnitude) and a minimum when it is parallel to the minor axis. If circularly polarized light passes through a crystal producing a relative retardation of an odd number of quarter wave lengths of a particular color the additional retardation between the components will cause the emergent light to be plane polarized in an azimuth which may be found by the analyzing Nicol prism. Such a crystal is called a quarter-wave plate. These plates can readily be prepared from thin sheets of mica.

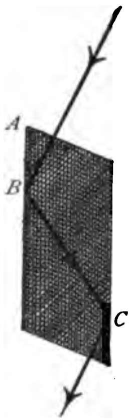


FIG. 349.

Another device for securing or testing circularly polarized light is Fresnel's rhomb (Fig. 349). A block of glass is cut with the angle at A equal to 54°, so that a pencil of light incident normally will be totally reflected at B and again at C, the angle of incidence being 54°. At each reflection at this particular angle a phase difference of an eighth of a period is introduced between the vibrations in and at right angles to the plane of incidence, and the emergent light is circularly polarized if the incident light is plane polarized at an angle of 45° with the plane of incidence. If this angle be other than 45° the amplitude of the two components will be different and the light will be elliptically polarized.

545. Rotation of the Plane of Polarization. If two Nicol prisms are set for extinction and a crystal of quartz cut with the face on which the light falls at right angles to the axis, or a solution of sugar or tartaric acid, is placed between them, the light will be restored. On turning the analyzer through a given angle depending on the thickness of the crystal or the solution, the light will again be extinguished. This shows that the plane of polarization has been rotated through this angle. Substances producing this effect are said to be naturally optically active.

If light passes through a quartz prism so cut that the light is transmitted in the direction of the optic axis it is found that there is a slight double refraction, so that spectral lines appear double. This shows that the two waves travel with slightly different velocities even along the optic axis; consequently the two wave surfaces cannot be tangent to each other (§ 528), but must be slightly

separated. This is not generally true of uniaxial crystals, but only of those which rotate the plane of polarization. It is found that the two waves are circularly polarized in opposite directions, so that this is a case of circular double refraction. As first suggested by Fresnel, it appears that when light travels along the optic axis of quartz it is divided into two circularly polarized components, which travel with different velocities. This offers a simple explanation of the rotation.

If the light incident on the quartz is plane polarized and vibrating in the direction AB (Fig. 350) there will be two circularly-polarized components transmitted, which on emergence will recombine into a plane polarized beam. If the velocities of propagation of the two components are the same the direction of polarization of the resultant will be unchanged. If one component travels faster than the other the plane of vibration will be rotated in the direction of rotation of that component which rotates through the greater angle a in traveling a given distance, *i. e.*, the component which has the smaller velocity of propagation. The two cases are illustrated in Figs. 350 and 351. If each circular displacement r and l is

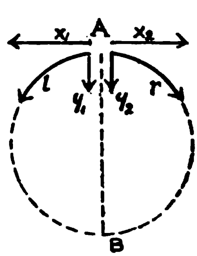


FIG. 350.

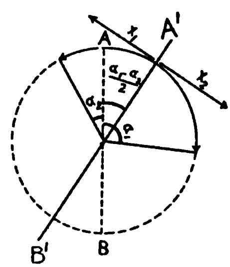


FIG. 351.

resolved into two linear displacements x and y (differing in phase by one fourth period), it is seen that in the first case the two x components at any point in the medium are equal and opposite, leaving the two y components in the same direction to combine in a plane polarized beam, the vibrations of which are in the same direction as those of the original beam. In the second case the x components and the y components are respectively unequal. If, however, we refer displacements to an axis of reference shifted through an angle $(\alpha_1 - \alpha_2)/2$ with the original direction of vibration it will be seen that with reference to this axis the x displacements will cancel each other. This line $A'B'$ then represents the final direction of vibration and the rotation is $(\alpha_1 - \alpha_2)/2$.

Some quartz crystals rotate the plane of polarization clock-wise looking in the direction of propagation, and are called right-handed; others produce rotation in the opposite direction, and are called left-handed. These two classes of crystals can be distinguished by inspection on account of certain unsymmetrical facets which are differently placed in the two cases. In order to overcome circular double refraction in quartz prisms used for spectroscopic purposes, the prism is made in two halves cut perpendicular to the axis, one of right and one of left-handed quartz. These are called Cornu prisms.

The rotation of the plane of polarization of light of the wave lengths corresponding to some Fraunhofer lines caused by a quartz plate of one mm. thickness is given below:

A	*B*	*C*	*D*	*F*	*G*	*K*
12.67°	15.75°	17.32°	21.70°	32.97°	42.60°	52.15°

As shown by these figures, the rotation varies very nearly inversely as the square of the wave length.

A number of other crystals produce rotation. Fused quartz shows no double refraction or rotation. These effects are evidently due rather to the crystalline arrangement of the molecules than to their individual structure.

A number of liquids, such as turpentine and the different sugars in solution, also cause rotation, that due to turpentine being left-handed, and that due to some sugars right-handed, of others left-handed. The vapors of such substances as turpentine also produce rotation. In such cases, as with quartz, there is circular double refraction, and the rotation is to be explained in the same way. In the case of liquids and vapors, however, the effect must be due to unsymmetrical structure of the molecule itself, as there is no crystalline structure, or if there is, the crystals are irregularly oriented. The amount of rotation varies inversely as the square of the wave length, and is proportional to the thickness of the medium, and also to the concentration in the case of solutions.

The specific rotatory power $[\alpha]$ is defined as being equal to the angular rotation produced by one decimeter of a substance of unit density (or one gram per cubic cm. in case of solutions). If

the observed rotation is a, $[a] = a/l\rho$, where l is the length and ρ the density, or if the percentage concentration is p, $[a] = 100a/lp$.

Saccharimetry. The rotatory power of the sugars is nearly proportional to the concentration, but is slightly affected by the presence of impurities. The percentage of sugar may be determined by measuring the rotation with a sensitive polariscope. This is called *saccharimetry*. Most sugars rotate to the right, but levulose rotates to the left. In some cases the specific rotatory power varies slightly with the concentration, and that of levulose is influenced by the temperature. The specific rotatory power for sodium light for some sugars at 20 degrees C. is given below (from Landolt, Optical Rotation). The positive sign indicates right-handed, the negative left-handed rotation, while p is the concentration coefficient.

Sucrose (cane sugar)	$+\ 66.44° + 0.0087p$
Dextrose	$+\ 52.50° + 0.0188p$
Levulose	$-\ 88.13° - 0.2583p$
Lactose (milk sugar)	$+\ 52.53°$
Maltose (malt sugar)	$+\ 140.4° - 0.0184p$

546. Rotation by Magnetic Field. Faraday discovered in 1845 that the plane of polarization of light passing through a refractive substance in a magnetic field is rotated if the light travels parallel to the force lines. No effect is produced by a magnetic field on light waves in free space, and in general the effect increases with the refractive power of the substance, being especially marked in dense flint glass and carbon bisulphide and very feeble in the case of gases. The rotation is usually proportional to the field intensity and to the thickness of the medium. Some substances cause right-handed and others left-handed rotation. The effect varies with the wave length. The rotation produced by one cm. thickness in a field of unit strength (Verdet's constant) is: For water, 0.0131°; carbon bisulphide, 0.0435°; dense flint glass, 0.06°. Kundt found that enormous rotations are produced by thin films of iron or other magnetic material in a strong magnetic field.

In naturally active substances the direction of rotation is independent of the direction of propagation of the light, so that if a rotated beam is reflected its plane is turned back to the original position. In magnetically active substances the direction of

rotation is reversed with reversal of the field, so that if the beam is reflected through the medium the rotation is doubled.

When a beam of plane-polarized light is **reflected** from a **metallic surface** a relative phase difference is introduced between components respectively in and at right angles to the plane of incidence, so that the reflected light is elliptically polarized, unless the incident light vibrates parallel or at right angles to the plane of incidence. Kerr found that if the light is reflected from the polished pole of an electromagnet it becomes slightly elliptically polarized, even under the conditions just mentioned.

547. Zeeman Effect. Faraday investigated the spectra of sources placed in a strong magnetic field, with a view of finding whether the radiation is affected thereby. His experiments were fruitless, but in 1896 Zeeman repeated them with a powerful Rowland grating, and discovered an effect which has proved to be exceedingly important in its bearing on theories of radiation. He placed a bunsen flame colored with sodium between the poles of an electromagnet. When the light from the source traveled either parallel or at right angles to the direction of the field, he observed a broadening of the spectral lines when the field was established. H. A. Lorentz pointed out that such effects were in harmony with an "electron" theory of radiation proposed by him, and predicted that further investigation would show the radiation to be polarized by the field, either circularly or plane, according to the direction in which it was viewed. Zeeman found this to be the case. In the simplest cases, when the light is viewed normally to the field, each spectral line is split into triplets, the vibrations in the central and undisplaced component being parallel to the force lines, those of the lateral and displaced components at right angles to the force lines. When the source is viewed parallel to the force lines single lines become doublets, the components being circularly polarized in opposite diections, and displaced on each side of the mean position of the line. In some cases the effects are much more complex, a large number of components being produced from single lines, but the simple case described above is fully explained by Lorentz's theory. He considers that the radiant center in the case of a line spectrum is an electron, a small electrically charged body, rotating within or about an atom in a circular or elliptical orbit. Other lines of evidence indicate that these electrons are held in their orbits by a positive charge on the atom, or rather "ion." These electrons communicate their motion to the ether. If an electron is rotating in an orbit parallel to the line of sight, only the component of motion at right angles to the line of sight is effective in sending waves to the eye, so that plane-polarized light is produced. If the orbit is at right angles to the line of sight, the electron stirs up an ether vortex, which reaches the eye as circularly-polarized light. A charge of electricity on a moving conductor is equivalent to an electric current (§ 623), and has a magnetic field which reacts with any external field. The electrodynamic

forces may easily be shown to accelerate component displacements of the charged body in one direction at right angles to the field, to retard those in the opposite direction, and not to affect displacements parallel to the field (§ 632). This explains the origin of the "triplets." In orbits at right angles to the field the speed of rotation will be accelerated when the electron is moving in one direction with respect to the field, retarded when moving in the opposite direction. When electrons are present moving in both directions two circularly-polarized rays, or ether vortices, are sent out parallel to the field.

No effect is produced on band spectra by a magnetic field. This is in harmony with the view that band spectra are characteristic of the undissociated neutral molecule, not subject to electrodynamic forces. A number of lines of evidence indicate that line spectra are possible only when the radiating vapor is ionized.

There is much evidence that the electrons are small negatively charged bodies, much smaller than even the hydrogen atom. Other examples of electrons are cathode rays and one class of radiation from radioactive bodies such as radium. (§§ 732, 747, 754.)

DISPERSION AND SELECTIVE REFLECTION.

548. Dispersion. It was pointed out in § 450 that dispersion due to refraction is irrational, that is, there is no simple relation

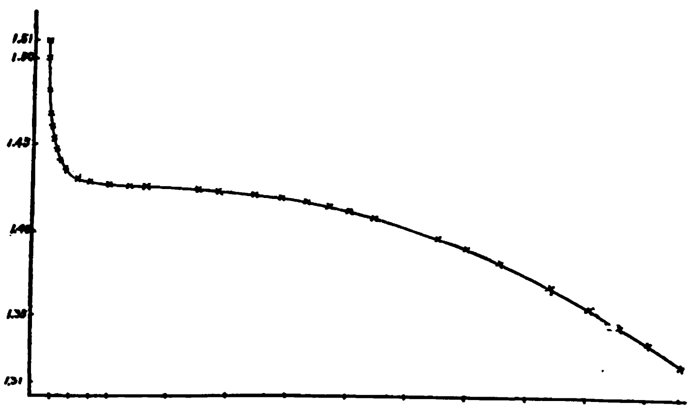


FIG. 352.

between the deviation of lines in the spectrum produced by a prism of the substance and the wave lengths, as there is in diffraction spectra. As a general rule the longer waves are less refracted

than the shorter, and the dispersion steadily diminishes in the direction of the longer waves, so that the red end of the spectrum is "telescoped" as compared with the violet. Within the limits of the visible spectrum the relation between the index of refraction and the wave length is closely expressed by the empirical relation (*Cauchy's* formula)

$$n = A + B/\lambda^2 + C/\lambda^4$$

where A, B, and C are constants varying with the substance. The dispersion curve of fluorite, showing the relation between index of refraction and wave length, is shown in Fig. 352.

549. Anomalous Dispersion. It is not always true that the deviation of waves by refraction increases as the waves become shorter. In 1860 Le Roux showed that iodine vapor transmits only the red and violet, and that the red is refracted more than the violet. In 1870 Christiansen found that in the case of fuchsine, an aniline dye, blue and violet are less refracted than red, the green is absorbed, and the other colors occur in the usual order. It has since been found that such anomalous dispersion is shown not only by a large number of substances such as the aniline dyes, but by the vapors of sodium and other metals, and, in fact, by almost every substance investigated in some part of its spectrum, visible or invisible. Anomalous dispersion always occurs in the neighborhood of what appears to be a strong absorption band, which is, more properly, a region where the light is selectively reflected rather than transmitted or absorbed. The index of refraction is abnormally increased on one side of this band and diminished on the other, resulting in the reversal of the corresponding colors in a spectrum formed by a prism of the substance. The dispersion curve of a substance between two regions of such selective reflection is shown in Fig. 353. Between these regions the curve resembles the normal dispersion curve shown in Fig. 352.

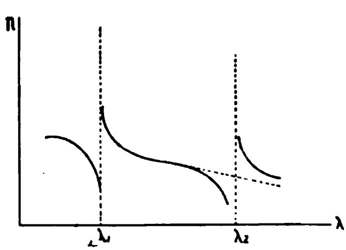


FIG. 353.

The absorptive power of substances showing anomalous disper-

sion is usually so great that it is impossible to secure a prism of sufficient angle to give a spectrum long enough to clearly show the effect. The method of crossed prisms is well adapted for showing it. If light passes in succession through two prisms with their refracting edges at right angles to each other the resultant spectrum will usually be a line or smooth curve inclined in direction to the edges of both prisms. If, however, one of the prisms

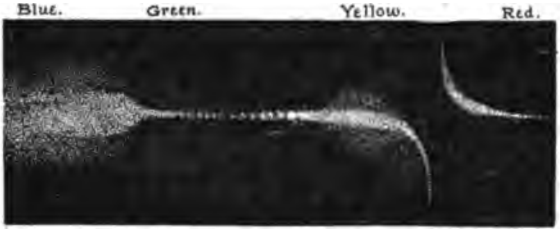


FIG. 354.

gives anomalous dispersion the resultant spectrum will be broken and irregular, as shown in Fig. 354, which illustrates the anomalous dispersion of sodium vapor in the neighborhood of the *D* lines, as photographed by *Wood.*

550. Selective Reflection. The color of natural objects is primarily due to selective absorption, the effective waves being those which escape absorption and become scattered. In the case of substances showing surface color, however, the effect is due to selective reflection, and the transmitted light is complementary to that reflected. This is the case with substances showing anomalous dispersion. The colors which are selectively reflected lie between the colors which are transmitted and anomalously dispersed. The so-called "absorption" bands so often referred to in this connection are thus seen to be largely due to lack of transmission because of reflection. As the reflecting power of the substance is abnormally high in such regions, they are said to show metallic reflection for the colors concerned. Recent investigations show that most substances exhibit anomalous dispersion in some region of their spectrum. For example, quartz, rock salt, and fluorite show anomalous dispersion and metallic reflection for certain very long waves. For radiation of wave length 611,000 Ångström units

(reflected from sylvite) quartz has an index of refraction of 2.12, considerably greater than that of the shortest ultra-violet waves.

551. Isolation of Long Waves. As all known substances are opaque to very long waves, and as grating spectra overlap in the region of long waves to such an extent as to make the method of dispersion by diffraction inapplicable, for a long time there seemed to be no way of isolating and studying such waves. In 1896 Rubens and E. F. Nichols showed that the property of selective reflection might be used. Suppose that a substance showing anomalous dispersion has a reflective power of 90 per cent. in the region of the wave lengths λ_1, λ_2, etc., while in other regions of the spectrum the reflective power is only 5 per cent. After three successive reflections 73 per cent. of the energy of the indicated wave lengths will remain, and practically no energy of the other wave lengths. Having thus isolated these particular waves, their length may be measured with a coarse diffraction grating, without the confusion due to overlapping of regions of shorter wave length. The lengths of some of the residual waves (Reststrahlen) obtained in this way are in Angström units

Quartz, 85,000, 90,200, 207,500. Fluorite, 237,000.
Rock salt, 512,000. Potassium chloride (sylvite), 611,000.

552. Theory of Anomalous Dispersion and Selective Reflection. It is believed that these effects are due to resonance, the free periods of the vibrating parts of the molecules being the same as that of the waves selectively reflected. According to the latest point of view, the vibrating element of the molecule is the electron. Selective reflection may be considered as the re-radiation of ether waves by the electrons, just as a tuning fork re-radiates sound waves after being excited by resonance. There is in such cases little "frictional" absorption of energy, which is completely transformed, to heat, not re-radiated. It may be shown from mechanical analogies and electrical theory that the rate of propagation of waves through a medium will be accelerated or retarded if the medium contains vibrating elements which have a free rate of vibration slightly greater or less than that of the waves.

A complete dispersion formula, taking account of regions having anomalous dispersion for wave lengths λ_1 and λ_2 is

$$n^2 = A + \frac{B}{\lambda^2 - \lambda_1^2} + \frac{C}{\lambda^2 - \lambda_2^2}$$

where λ_1 and λ_2 are the lengths of the light waves having the same rate of vibration as the electrons of the substance. This gives a discontinuity in n, the refractive index, and anomalous dispersion for these wave lengths.

The electron theory, first put on a definite basis by Zeeman's discovery, promises to give an explanation of radiation and most of the optical properties of bodies.

REFERENCES.

PRESTON'S *Theory of Light.*
WOOD'S *Physical Optics.*
EDSER'S *Light for Students*

These three books give a more advanced treatment of the subject than the ordinary text-book, but contain much interesting material which can be understood by beginners

TAIT'S *Light* also gives a somewhat advanced treatment.
TYNDALL'S *Light.*
STOKES' *Lectures on Light.*
HASTING'S *Light.*

These three books give a very interesting and simple popular account of the subject.

S. P. THOMPSON'S *Light, Visible and Invisible,* is an exceedingly interesting popular discussion of modern discoveries, not only in light, but in related phenomena, such as Röntgen rays and electric waves.

LE CONTE'S *Sight* treats of the eye and vision in a popular way.

ABNEY'S *Color Measurment and Mixture* and *Color Vision* give an elementary but complete discussion of the subject.

CHURCH'S Colour is another interesting little book on the same subject.

BALY'S *Spectroscopy* is an excellent presentation of spectroscopic methods and theory.

WATT'S *Introduction to Spectrum Analysis* is a somewhat more popular book than the above, and gives tables of wave lengths.

CLERK'S *Problems in Astrophysics* is a very interesting account of the uses of the spectroscope in astronomical work.

MICHELSON'S *Light Waves and their Use* gives an account of interferometers and their applications.

DERR'S *Photography* is an excellent elementary discussion of the photographic art.

The *Scientific Memoirs Series* contains the following reprints of important original papers, most of them presented in simple language: Prismatic and Diffraction Spectra, Fraunhofer; The Wave Theory of Light, Huyghens, Fresnel; Laws of Radiation and Absorption, Kirchhoff and Bunsen; The Effects of a Magnetic Field on Radiation, Faraday, Kerr, Zeeman.

PROBLEMS.

1. A man is 5 feet 10 inches high. What is the shortest plane mirror in which he can see his full-length image? Ans. 35 in.

Reflection.

2. Two plane mirrors are parallel to each other at a distance of 30 cm. Find the distance from each mirror of the three nearest images in each of an object between them and 10 cm. from one. Ans. 10, 50, 70; 20, 40, 80.

3. A beam of light is reflected from a plane mirror revolving clockwise about a vertical axis ten times per second, falls on a neighboring mirror revolving anticlockwise fifteen times per second, and then on a wall 10 meters away. With what speed does the spot of light cross the wall?

Ans. 3141.6 *m/sec.* anticlockwise.

4. A meter rod lies along the axis of a concave mirror of 20 cm. focal length, one end in contact with the mirror. Describe the images formed, and calculate the position of the first, fifth, tenth, twentieth, fortieth, and

one-hundredth cm. marks, and the length of each division at these points (assuming the rod to be 2 cm. wide).

Ans. Virtual distances, 1.05, 6.67, 20, ∞; real, 40, 25. Lengths, 2.1 2.67, 4, ∞; 2, 0.5.

5. Prove by graphical construction the statements made in § 441 concerning ellipsoidal, hyperboloidal, and paraboloidal mirrors.

6. Show by diagrams the successive shapes of the wave reflected from a hemispherical concave mirror as it passes from the mirror to a point beyond the focal cusp. (This surface must everywhere be normal to the "rays" which it cuts.)

7. A convex mirror has a focal length of 25 cm. Calculate the position and the height of the image of an object 10 cm. high and 15 cm. in front of the mirror. Ans. 9.4, 6.3.

8. A paper square with sides two cm. in length lies in and parallel to the axis of the above mirror at a distance of 40 cm. Describe the shape of the image, and calculate the lengths of its sides and the angles between them. Ans. Quadrilateral; sides normal to axis, 0.765, 0.744. Distance between them, 0.364; 91° 39′, 88° 21′

9. The sun has an angular magnitude of 32′. What is the size of the solar image formed by a concave mirror of 50 feet focal length? Ans. 5.58 in.

Refraction. **10.** A layer of ether ($n = 1.36$) 2 cm. deep floats on water ($n = 1.33$) 3 cm. deep. What is the apparent distance of the bottom of the vessel below the surface? Ans. 3.72.

11. An object is viewed through a cube of glass ($n = 1.55$) 10 cm. thick, in a direction at an angle of 60° with the normal to the glass surface. What is the lateral displacement of the image? Ans. 4.59.

Lenses. **12.** A convex lens 25 cm. from a candle-flame 5 cm. high forms an image of the latter on a screen. When the lens is moved 25 cm. further from the candle an image is again formed on the screen. Calculate the focal length of the lens, the distance of the screen from the candle, and the size of the two images.

Ans. 16.67; 75; 10, 2.5.

13. Show by graphical construction whether it is possible to construct a single thick double convex lens which will give a real erect image; and another which will give an inverted virtual image.

14. A candle flame 100 cm. from a convex lens of focal length 90 cm. is displaced 2 cm. away from the lens at the rate of 1 cm. per second. What is the displacement and the average velocity of its image?

Ans. 135 cm. toward lens; 67.5 $\frac{\text{cm.}}{\text{sec.}}$.

15. A convex lens ($n = 1.54$) has a focal length of 40 cm in air. What is the focal length in water ($n = 1.33$)? Ans. 136.8.

16. The images of objects seen through a spherical flask or cylindrical glass of uniform thickness are of diminished size. Explain.

17. Two convex lenses of focal lengths 20 and 30 cm. are 10 cm. apart. Calculate the position and length of the image of an object 2 cm. long 100 cm. in front of the first lens. (Consider the image due to the first lens to be the object for the second.)

Ans. 10 cm. beyond second lens; length, 0.33.

18. Replace the first lens in the above problem by a concave lens of the same focal length and determine the position and magnitude of the image.

Ans. 243 cm. to left of second lens; 3.04.

19. When focused on a star, the distance of the eyepiece of a telescope from the object lens is 50 cm. To see a certain terrestrial object clearly

the eyepiece must be drawn out 0.2 cm. What is the distance of the object from the observer? Ans. 125.5 m.

20. In the above example, if the eyepiece has a focal length of 1 cm., and if the object referred to is a tree 10 feet high, what is the size of the image formed by the object lens? What is the angular magnitude of the image formed by the eyepiece? Ans. 1.21 cm.; 62° 21′.

21. A double convex lens with faces having a radius of curvature of 30 cm. gives a real image at a distance of 60 cm. of an object 40 cm. away. What is its focal length? Its index of refraction? Ans. 24; 1.625.

22. An achromatic lens is to be made of a combination of a crown glass double convex lens ($n_D = 1.51$, $n_F = 1.52$) and a plano-concave flint glass lens ($n_D = 1.65$, $n_F = 1.66$), the adjacent faces to fit together and the focal length to be 50 cm. Calculate the radii of curvature of the faces. Ans. $r_1 = \infty$; $r_2 = r_3 = 7$; $r_4 = \infty$.

Photometry. **23.** A candle is placed 10 cm. in front of a concave mirror of 20 cm. focal length (assumed to be a perfect reflector). What is the illumination on a screen 100 cm. from the candle along the mirror axis, as compared with (I), that due to the candle alone? Ans. 3.31 I.

24. Solve the above problem after substituting a convex mirror of the same focal length for the concave mirror. Ans. 1.33 I.

25. Two sources have candle power 16 and 97 respectively. At what point between them must a screen be placed to be equally illuminated by the two? Ans. 0.288 l from fainter source.

26. Foucault in his arrangement for measuring the velocity of light (§ 410) placed the lens between the source and the revolving mirror. Show that with this arrangement—(a) The stationary mirror must be concave, with center of curvature at the axis of the revolving mirror; (b) that the stationary mirror cannot be placed far away from the revolving mirror unless its aperture is correspondingly enlarged, if the reflected beam is to have sufficient intensity. Show that Michelson's arrangement obviates these disadvantages.

Dispersion. **27.** A 60° prism has an index of refraction of 1.62 for the D lines and 1.63 for the F line. If white light is incident at an angle of 45 degrees, what are the respective angles of emergence for these two colors? Ans. 65° 19′.8; 66° 40′.8.

28. In the above case, what is the angle of minimum deviation for each color? If the spectrometer telescope has a focal length of 30 cm., what is the length of the spectrum between D and F when the prism is set for minimum deviation for the D lines? Ans. D, 48° 11′.2: F, 49° 10′4; 7.1 cm.

Total Reflection. **29.** Looking down into a cylindrical drinking glass partly filled with water, one cannot see external objects through the sides of the glass, but if a finger is firmly pressed against the side of the glass it can be seen from above. Explain.

30. Light incident internally on the surface of a glass prism at an angle of 56° is totally reflected from a drop of liquid in contact with the glass. If the index of refraction of the latter is 1.62 for sodium light, what is the index of refraction of the liquid? Ans. 1.343.

Interference. **31.** In a system of Newton's rings due to a convex lens resting on a plane surface the 25th ring is one cm. from the center, when sodium light is used. What is the thickness of the air film at that point, and what is the radius of curvature of the lens? Ans 0.00751 mm.; 6.67 meters.

32. If the air film is replaced by water in the above example, what will be the distance of the 25th ring from the center? Ans. 0.97 cm.

33. Light from a narrow slit passes through two parallel slits 0.2 mm. apart. The interference bands on a screen 100 cm. away are 2.95 mm. apart. What is the wave length of the light? Ans. 0.00059 mm.

34. The angles of a Fresnel biprism are 10′ and the index of refraction 1.62. What is the distance between the two images of a slit 20 cm. from the prism? What is the width of the interference bands of sodium light formed on a screen 50 cm. beyond the prism? What is their width if light of the wave length of the F line is used?

Ans. 0.724 mm.; 0.57 mm.; 0.47 mm.

35. A film of glass of index of refraction 1.54 is introduced in one of the interfering beams of a Michelson interferometer, and causes a displacement of 20 fringes of sodium light across the field. What is the thickness of the film? Ans. 0.0218 mm.

36. The D lines in the spectrum of the second order formed by a Rowland concave grating of 15 feet radius of curvature are 315 cm. from the slit. What is the distance between rulings? Ans. 0.00171 mm.

Diffraction. **37.** The central maximum of the diffraction bands of sodium light produced by a narrow slit on a screen at a distance of 100 cm. is 2 mm. wide. How wide are the other maxima and the slit? Ans. 0.589 mm.; 1 mm.

38. Explain the diffraction bands in the shadow of a needle or wire (Fig. 301).

39. Describe and explain the appearance of the filament of a distant electric light as seen through very small pinholes of different sizes.

40. Two narrow slits 0.1 mm. apart are illuminated by sodium light. What must be the diameter of a lens 5 meters away to clearly resolve the images of the two slits? Ans. 2.95 cm.

41. In the above case, at what distance will the same lens clearly resolve the images of the slits if they are illuminated by light of wave length corresponding to that of the F line? Ans. 6.05 m.

Polarization. **42.** Plane polarized light falls normally on a plate of quartz with faces parallel to the axis. If the vibrations of the incident light are at an angle of 30° with the principal plane, calculate the relative intensities of the transmitted ordinary and extraordinary rays. Ans. 0.25, 0.75.

43. In the above case, if the crystal is 1 mm. thick, what is the difference of phase upon emergence of the ordinary and extraordinary rays of sodium light (§ 536). Ans. 0.45λ.

44. A crystal of Iceland spar cut with faces parallel to the axis is two cm. thick. How far below the upper surface are the ordinary and extraordinary images of a pencil mark on the lower face? Ans. 1.206, 1.346.

45. Through how many degrees will a column 20 cm. long of a 10 per cent. solution of cane sugar rotate the plane of polarization of sodium light?

Ans. 132°.88.

Spectrum. **46.** On mapping the spectral intensity curve of an incandescent source it is found that the maximum intensity is at a wave length 12,000 Ångström units. What is the temperature of the source? Ans. 2399° abs.

47. The displacement of the F line of hydrogen (wave length 4861 Ångström units) in the spectrum of a star is .1 of a unit toward the violet. What are the direction of motion and the velocity of the star in the of the source? Ans. 6.2 $\frac{\text{k}}{\text{sec.}}$ toward earth.

ELECTRICITY AND MAGNETISM.

By Arthur W. Goodspeed, Ph.D.,

Professor of Physics in the University of Pennsylvania.

MAGNETS AND MAGNETIC FIELDS.

553. Magnets and Magnetism. The terms **magnet** and **magnetism** were derived from the name of the city of Magnesia in Asia Minor near which was first found an iron-ore having the special property of attracting iron. This mineral has received the name *magnetite.* It was known as early as the tenth or twelfth century that a piece suspended by a thread always sets itself in the same position relatively to the north and south line. Because of this directive property the substance was also called *lodestone.* Little more was observed of the properties of this mineral till the appearance of "De Magnete" by Dr. Gilbert in 1600. He located centers of attraction which he called *poles* of the magnet, and he called the straight line joining the poles the *magnetic axis.* Gilbert also discovered the *field of force* surrounding a magnet, and by means of iron filings he studied the shape and properties of the field somewhat in detail. That cobalt and nickel are also attracted by a magnet was not discovered till nearly one hundred and fifty years after Gilbert's time, the former by Brandt and the latter by Cronstedt. In 1778 Brugmans discovered that bismuth is actually repelled by a magnet, while Faraday in 1845 found the same to be true of several other substances, as phosphorus, antimony, and copper.

554. Magnetic Substances. It appears then that substances are divided into two groups, those which when free behave like magnetite and tend to set their axes nearly parallel to the north and south line, and those which tend to place their longer dimensions across or perpendicular to this direction. Substances belonging to the former class are called **paramagnetic,** while those of the latter are called **diamagnetic** substances. Paramagnetic substances

are attracted by a magnet, while diamagnetic ones are repelled. The former group comprises only a few materials, notably iron, nickel, cobalt, manganese, chromium, and cerium. Of these, iron exhibits far the strongest effects. To the diamagnetic group belongs a much longer list of substances including bismuth, antimony, zinc, tin, cadmium, mercury, lead, silver, copper, gold, arsenic, uranium, rhodium, iridium, and tungsten.

The medium surrounding a body affects somewhat its magnetic action. A body ordinarily paramagnetic appears diamagnetic when immersed in a liquid more paramagnetic than itself.

Certain alloys of the non-magnetic metals, manganese, aluminum and copper, known as Hensler's alloys, show striking magnetic properties. These properties vary considerably with the previous thermal treatment of the specimen.

555. Magnetization. Magnets may be either natural magnets such as pieces of lodestone, or artificial ones, as pieces of iron or steel which have been made to take on the properties of a magnet. A piece of iron or steel when brought near to one of the poles of

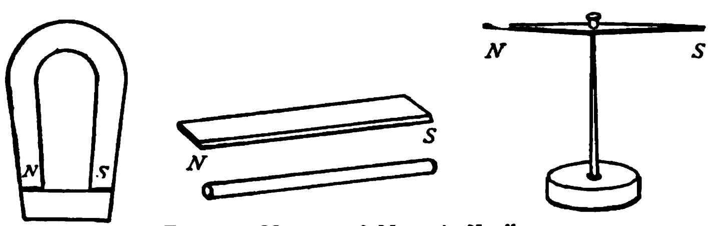


FIG. 355. Magnet and Magnetic Needle.

a magnet is found to possess the properties of a magnet. In the case of soft iron these properties nearly disappear when it is removed from the magnet, but in the case of the steel the magnetism is more or less permanent, depending on the quality of the specimen. Hard steel is difficult to magnetize but it retains its magnetism indefinitely. This power of retaining magnetism when not under the influence of a magnet is called *retentivity*.

Artificial magnets are prepared in various ways. They are usually made of steel or chilled iron and the harder the material the more difficult it is to make a strong magnet. When once magnetized, however, the harder the bar the more permanent is

the result. If a steel bar be rubbed lightly from end to end, always in the same direction, by the pole of a strong magnet, the bar will become magnetized. By repetition of the process the strength of the magnet is increased for a time until finally a condition is reached when no further rubbing has any effect. (A more effective method depending on the use of a current of electricity will be explained later in § 633.)

A heavy, thick bar is much more difficult to magnetize than a thin one. Hence it is customary when a large magnet is wanted to magnetize a number of thin pieces and then to bolt them together into one mass. In this way a *compound magnet* is made which is far more powerful than it would be possible to make from one piece. It must be noted that the resulting strength of a compound magnet is not by any means the sum of the strengths of its parts. This is because of the weakening effect of like poles on each other.

It is often convenient to be able to make use of both poles at the same time and this is made possible by bending the bar into the shape of a U or of a horseshoe.

556. Effect of Temperature. Gilbert states that iron at a red heat cannot be magnetized. In more recent times it has been shown by Hopkinson, Ledeboer, Rowland and others that up to about 680° C. pure iron does not change appreciably in its magnetic properties; that above this a rapid change begins, and at a temperature of about 870° its magnetic properties disappear altogether. For nickel the temperature of loss of magnetic properties is about 320°; for cobalt it is about 1100°.

557. Dual Nature of Magnetism. If a magnetized steel bar be examined along its length by hanging on soft iron nails of various sizes, it will be found that the heaviest nails will be supported at or very near the ends or poles and that from the poles towards the middle the strength decreases to zero. If the bar be poised or suspended so as to be free to turn in a horizontal plane, it will direct itself approximately north and south, and it has become customary to distinguish the end directed towards the north by the mark + or N and to call it the *positive* or *north-seeking* pole. The other end is called *negative,* and is distinguished by the mark — or S.

If two suspended magnets be brought near each other, it will

be noted that the two positive poles will repel each other, and also that the two negative poles will repel each other. A positive pole and a negative one will, however, attract each other. From these observations we may generalize that: *like magnetic poles repel while unlike magnetic poles attract each other.* Furthermore these forces may be greater or less depending on the strength of the poles, and on the distance between them. Two poles are said to be of equal strength when, if placed at equal distances from a third pole, they attract or repel it with equal forces. *A unit magnetic pole, or a pole of unit strength, may be defined as one capable of exerting a force of one dyne on another pole of the same strength at the distance of one centimeter in air.* Hence a pole of strength m is one that will exert a force of m dynes on a unit pole placed at unit distance from the former.

558. Coulomb's Law. Although we cannot isolate a single pole, yet in the case of a long thin magnet the space near one of its poles is practically controlled by that pole alone, the other being relatively so far away. The laws of magnetic attractions or repulsions were first investigated by Coulomb in 1785.

He showed that the two poles of a magnet are of equal strength and he investigated the law of variation of force with distance in the case of two poles of different magnets. This was done with his torsion balance. A long magnetic bar was suspended horizontally by means of a wire, the upper end of which was fastened at the middle of a circular graduated disk capable of being rotated a measurable amount. Another long magnet was fixed with its axis vertical and with one of its poles horizontally opposite to a pole of the suspended magnet so as to rotate the latter. The mutual force between the two at different distances was determined by measuring the angle necessary to turn the disc to keep the poles at a fixed distance apart. Magnets of different strengths were also used. The results of these experiments led to the two following generalizations:

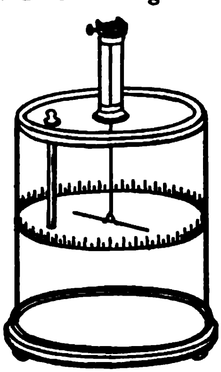


FIG. 356.

1st. The mutual force between two magnetic poles at a given

distance is directly proportional to the product of the strengths of the two poles.

2d. For the same poles the force is inversely proportional to the square of the distance between them.

The following expression combines both of these and is known as Coulomb's Law:

$$F = \frac{1}{\mu} \frac{mm'}{r^2}$$

in which F denotes the force, m and m' the strengths of the two poles and r the distance between them, all measured in absolute units. Coulomb's experiments were performed in air. Since his time it has been found that the force between two poles depends on the medium in which they are placed. μ is a constant depending on the medium through which the force acts and in the case of air it is necessarily unity (§ 565). The force may be one of attraction or of repulsion. Conventionally a repulsive force is assumed to be positive because it tends to increase the distance r.

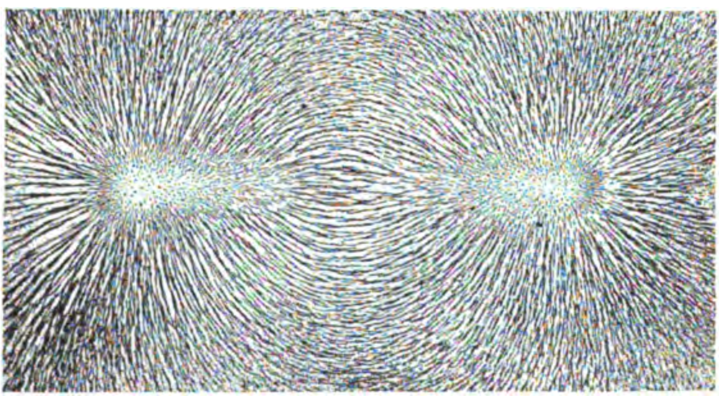


FIG. 357.

559. Magnetic Field. Since there is a mutual action between two magnets even when quite a distance apart, the space around a magnet is possessed of certain properties distinguishing it from other space. We express this by saying that a magnet creates around itself a field of force which we call a *magnetic field*. We

may define the *intensity* of a magnetic field at any point as equal to *the force that would be exerted on a unit N magnetic pole placed*

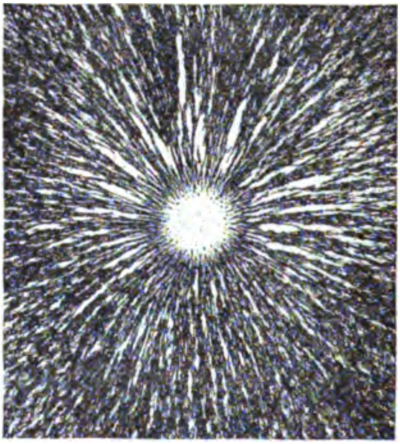


FIG. 358.

at the point. The character of such a field may be made clear to the eye and studied by placing a magnet in a horizontal position, covering it by a piece of card or of thin glass and sifting uni-

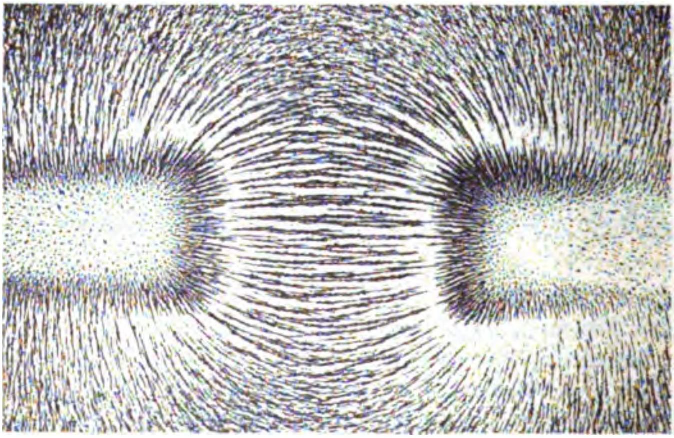


FIG. 359.

formly over the upper surface a layer of soft-iron filings. On tapping the disc the filings readily arrange themselves longitudinally according to a definite scheme depending on the shape of the magnet, the position of its poles, etc. Figs. 357–360 represent magnetic fields in special well defined cases. In Fig. 357 the field is produced by a single straight bar magnet with the poles near the ends. Fig. 358 shows the field looking end-on towards a bar magnet. Fig. 359 represents the field between the poles of two such magnets with unlike poles opposed. Fig. 360 shows the field between two like poles. The arrangement of the filings at once suggests the linear property of a field of force, and in fact it was suggested by Faraday to represent fields of force by lines of force, *the direction of the line at any point indicating the direction in which the resultant force acts,* and the closeness of

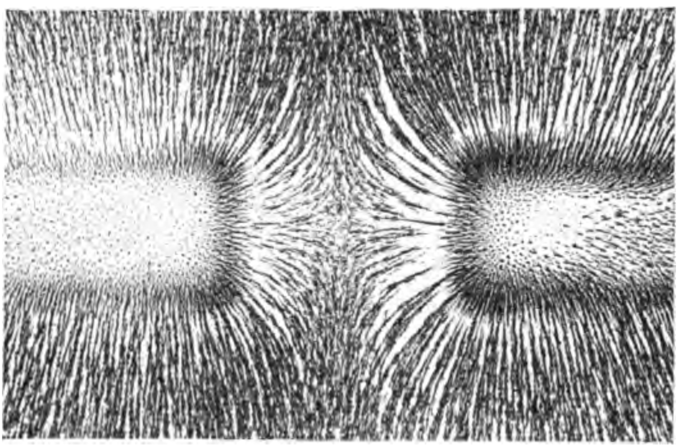


FIG. 360.

the lines indicating the intensity of the force. Following out this suggestion it is agreed now to represent graphically the intensity of the field by drawing lines of force, so that the number of lines per square centimeter perpendicular to their direction shall equal the intensity of the field. (Such lines might more properly be called *lines of intensity.*) If an area not unity be considered and if this be denoted by a, then the product of a by the average in-

tensity of the field equals the total number of lines of force through a and is called the ***magnetic flux*** through the area.

It is believed that all actions which seem to take place at a distance are really due to a strained state of the intervening medium. This strain was illustrated by Faraday, in the case of magnetism, by supposing the lines of force to be like stretched elastic threads, tending to contract and hence to pull in the direction of their length, and also to be mutually repellent in a direction perpendicular to their length. Another glance at Figs. 359, 360 will prove very suggestive of this idea. It is agreed to denote as the positive direction of a line of force that in which a positive magnetic pole tends to move. From this it follows that the lines of a single magnet emanate from the positive pole of the magnet and pass in at the negative pole.

If a field is such that the intensity at all points has the same value it is said to be a uniform field. Evidently the lines of force of such a field are parallel straight lines uniformly distributed.

560. Positions of Poles in a Magnet. Thus far we have not needed to define very precisely the position in a magnet of the points called poles. In fact it is evident from the figures in § 559 that the magnetic properties of a magnet are not confined to two definite points. We shall now for greater clearness define the position of a pole in a manner analogous to that used in defining the center of gravity of a body (§ 100). Suppose the magnet to be placed in a uniform magnetic field. The effect of the field on one end-portion of the magnet consists of a system of parallel forces exerted on the parts of the iron. The resultant of these forces passes through a definite point in the magnet which is defined as the position of the pole. Thus a pole of a magnet is analogous to the center of gravity of a body. The analogy is in fact still closer. For, as we cannot define any point as the center of gravity of a body unless we assume the forces of gravity to be parallel, so we cannot strictly define any point as the pole of a magnet, when the end of the magnetized bar is in a field the lines of force of which are not parallel. The distance apart of the two poles of a long magnet is about 0.85 of the length of the bar, but it varies with the dimensions of the bar.

561. Field Due to a Unit Pole. Evidently the field due to a

single pole is represented by lines of force radiating from the pole. If, with a unit pole as a center, we draw a sphere of one centimeter radius the area of the surface will be 4π sq. cm. As the intensity of the field over this surface is unity, there must be one line for each sq. cm. or 4π lines in all radiating from the unit pole. In general then from any pole of strength m, $4\pi m$ lines emanate, and the total flux through any surface surrounding the pole is 4π times the strength of the magnetic pole. (Unless otherwise stated the surrounding medium will always be assumed to be air.)

In the above we have supposed lines of force drawn so as to represent the intensity at unit distance from the pole. Do these same lines represent the intensity at any other distance? To answer this suppose a sphere of radius r drawn with the pole as center. The intensity at a point on its surface is m/r^2 (§ 558). The flux of the above lines through it is $4\pi m$ and its area is $4\pi r^2$; hence the number of lines per unit area through it is m/r^2, that is, equal numerically to the intensity. Hence lines of force drawn to represent the intensity at any point also represent by their density the intensity at any other point along their course. This can be proven to be true of a field due to *any* distribution of magnets.

562. Magnetic Moment. If a short bar magnet be placed in a magnetic field, it tends to assume a position with its axis parallel to the lines of force in the field. In other words when in an oblique position, as shown in Fig. 361, the poles tend to move along the lines of force in opposite directions, and a couple is developed which, if the bar be free to turn, rotates it till its axis is parallel to the field. In this latter position the moment of the couple becomes zero. If the strength of a pole be m, and that of the field be H, the force exerted on the pole is mH dynes. The position of the magnet in which the rotating couple is greatest is evidently perpendicular to the field. In this position the value of the moment of the couple is mHl, since each pole has half the length of the magnet as the arm of its moment. The product ml, or the strength of either pole multiplied by the

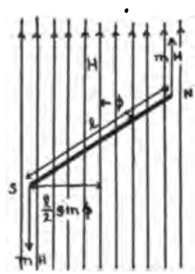


FIG. 361.

distance between the poles is called the ***magnetic moment M*** of the magnet. If the angle between the magnetic axis and the field be ϕ, the restoring couple is $mH \cdot l \sin \phi$ or $MH \sin \phi$.

In the above we have for simplicity defined magnetic moment as the product of pole strength and the distance between the poles. This definition is, however, not quite satisfactory, since both m and l are somewhat indefinite (§ 560) and we may now replace it by a better one. The magnetic moment of a magnet equals the couple that acts on it when it is placed in a uniform field of unit intensity and at right angles to the lines of force. In all practical work it is with the magnetic moment of a magnet (and quantities defined from it) that we have to do (§ 564) and by the above definition we practically get rid of the consideration of pole strength and distance between the poles.

563. Effect of Dividing a Magnet. Molecular Hypothesis. If a long bar magnet be divided in the middle and the parts tested, it will be found that each is a complete magnet, the ends originally together at the middle having opposite polarities and each opposite to that at the other end of the new magnet to which it belongs. If now the two pieces be again adjusted together, the combination will behave just as before dividing, the two opposite poles in the middle neutralizing each other as far as any outside effect is concerned. Each of the pieces due to the first division may again be made into two complete magnets, and so on indefinitely. From this we are led to the idea that perhaps this subdivision might have gone on even to the molecules of the substance, and that really each molecule of the magnet is itself a little natural magnet. Thus it would follow that, in an unmagnetized bar, the little molecular magnets would be situated with their axes in all possible directions, so that the resultant effect at any point would be zero. In the act of magnetizing such a bar, however, many of the molecules would be forced to rotate so as to present their positive poles all in the same direction. This would cause the corresponding end of the bar to be positive also, while the other end would be negative. On the foregoing hypothesis the detailed behavior of magnets can be satisfactorily explained. For example, bars of both iron and steel when placed in a magnetic field become magnets. This is because, as already stated, a magnet tends to place its axis parallel to the lines of force in a field. The molecules do this both in the case of the iron and of the steel, more easily in the former. When the bars are removed from the field, however, the iron molecules swing back toward their old positions. the bar losing thereby a part of its magnetism, while the steel bar retains to a great extent the new arrangement and therefore becomes a permanent magnet.

564. Measurement of MH. We have seen (§ 562) that a magnet of moment M when placed in a field of intensity H is acted on by a couple whose moment is $MH \sin \phi$, if it makes an angle ϕ with the direction of the field. If it be free to swing, its angular acceleration will be (§ 82)

$$\alpha = -\frac{MH \sin \phi}{I}$$

where I is the moment of inertia of the magnet about the axis of rotation (§ 84). For small angles the angle may be substituted for the sine, and the expression for α becomes proportional to the angular displacement. The motion is then harmonic (§ 118), and

$$T = 2\pi \sqrt{-\frac{\phi}{\alpha}}$$

Hence

$$T = 2\pi \sqrt{\frac{I}{MH}}, \quad \text{or} \quad MH = \frac{4\pi^2 I}{T^2}$$

This formula for the period is identical with that relating to the compound pendulum (§ 120), in which g, the intensity of the gravitation field, corresponds to H, the magnetic intensity, and mh for the pendulum corresponds to ml ($= M$), the magnetic moment of the magnet.

Since, from the above,

$$H = \frac{4\pi^2 I}{M} \cdot \frac{1}{T^2}$$

fields of different strengths may easily be compared by determining T in each case with the same magnet, the value $4\pi^2 I/M$ being constant.

We have then

$$\frac{H_1}{H_2} = \frac{T_2^2}{T_1^2}$$

Measurement of M/H.—If the magnet used in the last section now be placed with its axis at right angles to the field H, its own field at some dis-

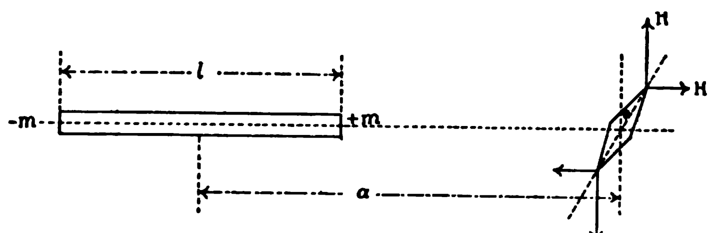


FIG. 362.

tance along the axis will be nearly uniform within narrow limits and will be perpendicular to H. Let its value be H' at a point where is suspended a very

small magnet that can rotate about an axis perpendicular to both fields. Fig. 362 will show the arrangement, the small magnet having turned through an angle ϕ till the moments of H and of H' are balanced. We have then, $\tan\phi = H'/H$. Now H' results from the effect of $+m$ at the distance $(a - l/2)$, and of $-m$ at the distance $(a + l/2)$.

Therefore

$$H' = \frac{m}{\left(a - \frac{l}{2}\right)^2} - \frac{m}{\left(a + \frac{l}{2}\right)^2}.$$

On reducing and writing M for ml, we get

$$H' = \frac{2Ma}{\left(a^2 - \frac{l^2}{4}\right)^2}$$

As l is small when compared with a, $l^2/4$ may usually be neglected in the above expression, and it then reduces to $H' = 2M/a^3$. Substituting this in the formula for $\tan\phi$ above, we get $\tan\phi = 2M/a^3H$, and

$$\frac{M}{H} = \tfrac{1}{2}a^3 \tan\phi$$

Since from the last expression

$$H = \frac{2M}{a^3 \tan\phi}$$

we can compare the field intensities at two places by using the same apparatus at the two places, in which case $2M/a^3$ will be constant and

$$\frac{H_1}{H_2} = \frac{\tan\phi_2}{\tan\phi_1}$$

Determination of M and H.—From the expressions just derived for both the product and the ratio of M and H it follows that

$$M = \frac{\pi}{T}\sqrt{2Ia^3 \tan\phi}, \quad \text{and} \quad H = \frac{2\pi}{T}\sqrt{\frac{2I}{a^3 \tan\phi}}$$

in which all the quantities on the right are observed from experiment.

565. Magnetic Induction. If a piece of any paramagnetic substance, notably soft iron, be brought near either pole of a magnet, or, in other words, if it be placed in a magnetic field, it becomes temporarily a magnet with the poles and all the other properties of a magnet. This effect on such a piece of iron is called "*induction*," and may be shown by the use of iron filings as in Fig. 363. The lines of force are made to change their direction and to enter the surface of the iron, causing therein a much stronger field than was present before.

Consider a very long bar of iron placed in a uniform magnetic field. Poles are developed at its ends and these also produce a magnetic field but, if the bar is long enough, their effect near the

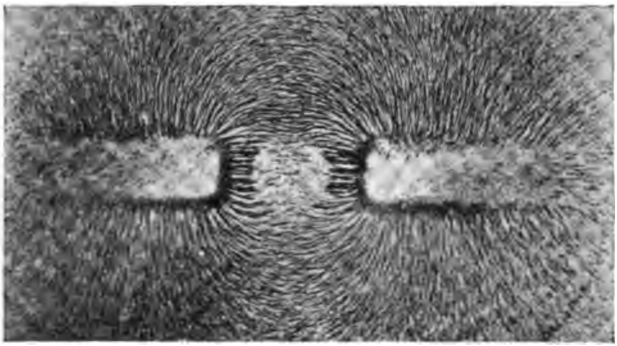


FIG. 363.

middle of the bar may be neglected. We shall consider the state of things near the middle of the bar.

Suppose the intensity of the field, assumed uniform before the presence of the iron, was H, and that through the iron the flux density has been increased to a value B, then $B = \mu H$, in which B is called the ***magnetic induction***, and μ the ***permeability***. Let us now suppose a very narrow cut to be made across the middle of the bar. On one side of the cut a positive pole will appear and on the other side a negative pole. If m be the strength of each of these poles and a its cross-section, then the ratio of m to a is called the ***intensity of magnetization***, or $I = m/a$. The ratio of I to H is called the ***magnetic susceptibility*** and is generally denoted by k. We have seen also that the flux emanating from m is $4\pi m$ and is outward from a positive pole, inward towards a negative pole. The total flux through the bar is then made up of two parts, viz., Ha due to the original field and $4\pi m$ or $4\pi Ia$ due to the induced magnetism in the bar, or $\Phi = Ha + 4\pi Ia$. Since by definition $B = \Phi/a$, it follows that

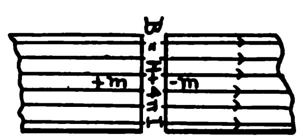


FIG. 364.

$$B = H + 4\pi I$$

$4\pi I$ is then the increase in the induction due to the presence of the iron. (Strictly speaking the lines of force H in air give rise to lines of induction in air but the two sets of lines are equal in number, since for air μ has been taken as unity.) Since $I = kH$, and $B = \mu H$, we get by substituting in the above formula,

$$\mu = 1 + 4\pi k$$

In the study of these magnetic properties it is much more convenient to produce and control the magnetic fields by means of the magnetizing effect of currents of electricity (623). By taking a specimen in the form of a ring and magnetizing by a helix, the difficulty of free ends is avoided.

566. Relations of I and B to H. When an iron rod is placed in a magnetic field, H, of increasing strength, the value of I increases also for a time but finally approaches a limit, called the *saturation value,* depending on the particular kind of iron. The curves (1) and (2), Fig. 365, show the relation of I and H for

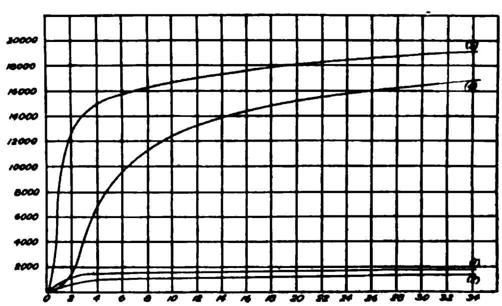


FIG. 365.

average specimens of soft iron and of cast steel. Similarly the relation between B and H is shown in the curves (3) and (4), (3) for soft iron and (4) for cast steel. Since I in the equation $B = H + 4\pi I$ approaches a limiting value, the curves relating B and H will approach a straight line ($B = H + 4\pi I'$) for large values of H. The following table indicates, for certain definite values of H, the corresponding values for B and for μ in the cases of certain samples of three kinds of iron. These results were taken from the work of Hopkinson and of M. E. Thompson.

	Wrought Iron.		*Cast Iron.*		*Cast Steel.*	
H	B	μ	B	μ	B	μ
10	12,400	1,240	5,000	500	9,800	980
30	15,100	503	7,400	247	13,500	450
50	15,950	319	8,450	169	14,700	294
70	16,500	236	9,200	131	15,330	219

567. Hysteresis. Suppose a piece of neutral iron to be placed in a region of zero magnetic intensity so that $H = 0$ and let H gradually increase causing thereby magnetization of the iron, and correspondingly increasing values of B. It has been found by experiment that the curve representing the relation of H to B is of the form OP, Fig. 366. Let H now gradually be brought back to zero; it is found that the return BH curve is not at all coincident with OP, and that B does not return to zero. This means, of course, that the iron does not lose all its magnetism, even though the magnetizing field has been brought back to zero. The induction is now equal to B_1 and is represented by Ob_1 on the diagram. B_1 is called the *retentivity* of the specimen. In order completely to demagnetize the iron or bring B back to zero, H must be reversed in direction and attain a certain negative value, H_1, represented by $-Oh_1$. If now the negative magnetizing field be increased so that its numerical value is equal to $-H$ and represented by $-Oh$, the induction will have the value $-B$ and be represented by $-Ob$. A reversal again of the magnetizing field bringing it back to zero brings the induction back to a value $-B_1$ represented by $-Ob_1$, and if the field continue to increase to Oh_1 and to Oh, the induction will increase first to zero and finally to B, represented by Ob as before. The field intensity represented by $-Oh_1$ is called the *coercivity* of the iron. In hard steel its value is of the order of magnitude of 100 C. G. S. units; in soft steel it is about a fifth

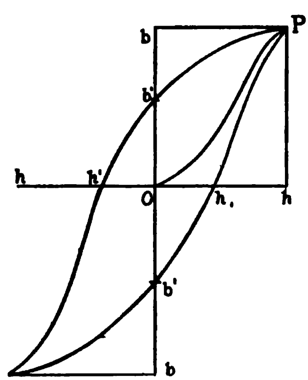


FIG. 366.

as great, and in soft iron it may be less than 2 units. Since for decreasing values of H the value of B is always greater than for the same value of H when the latter is increasing, B always lags behind H in its variation. This phenomenon is called *hysteresis* by Ewing who did much valuable work in this subject. In many forms of electromagnetic machinery the rapid magnetization in alternate directions is an essential feature of its operation. Such reversals are always accompanied by a production of heat, that is, by a dissipation of useful energy. For this reason an accurate knowledge of the hysteresis properties of all the iron and steel used in the construction of such machinery is most important. The hysteresis cycle is the broader the greater the coercivity.

Ewing attributes the opposition to the magnetization and demagnetization of a body to the mutual action of the molecular magnets composing the body. He constructed a model of a magnetic body out of a large number of small magnets mounted on pivots and free to turn. The influence of a weak field on such a system was found to be a slight rotation of many of the small magnets from their usual positions. The usual positions were resumed, however, as soon as the field was removed. A stronger field turned the magnets more and more, up to a critical limit beyond which an entirely new stable arrangement was assumed by the magnets. On weakening the field gradually, the new configuration tended to persist and was not broken up until a sufficiently strong field in the opposite direction was applied. In fact such a model exhibits all the stages of hysteresis explained above.

It may be shown that the formula for the work done in magnetizing a rod of magnetic material is

$$W = \frac{la}{4\pi}\int H dB$$

in which a is the cross-section and l the length of the bar made great enough that the effect of the poles on the value of H may be neglected. If the specimen is in the form of a ring there will be no free poles and l may be taken as the mean circumference. This integral evidently equals the area of the diagram (allowance being made for the scale of drawing); and for a complete cycle the integral therefore equals the area, A, of the hysteresis loop. Hence, since la is the volume of the specimen, the heat produced per cycle is $A/4\pi$ expressed in mechanical units (*i. e.*, in ergs if H and B are in absolute units).

A proof of this formula may be found in technical books on the subject.

568. Diagrammatic Properties. It has already been noted (§ 554) that the great majority of substances are diagrammatic which means that their permeabilities are less than one or that of air. Diamagnetism, however, is far less pronounced than paramagnetism. For example bismuth, the most diamagnetic substance known, has a permeability of 0.9998, while iron may have a permeability as high as 2000.

Since in diamagnetic bodies the ratio of B to H is less than unity it is evident that the induction is less than the field when the body is absent, and therefore the lines or tubes of induction may be thought of as being in the opposite direction to that of H and so to neutralize H somewhat.

TERRESTRIAL MAGNETISM.

569. The Earth's Field. We have seen that in a magnetic field a small suspended magnet will set its axis parallel to the lines of force, and, since the very first observers of magnetic properties noted that a freely suspended bar places itself nearly north and south, it follows that a magnetic field must surround the earth. The directive influence of this field on a magnetic needle was known by the Chinese nearly three thousand years ago and, according to Humboldt, they used a floating magnetized needle for the purpose of navigating the Indian Ocean as early as the third or fourth century. In the more modern mariner's compass a circular card, graduated on the edge into 32 equal parts called *points,* is fastened to the needle and rotates with it, the magnetic axis of the needle passing through the center of the card and through a pair of opposite points marked N and S. The relative position of the needle and the direction in which the bow of a boat points at any time may be readily observed. In order that the instrument may remain as nearly as possible in a horizontal position during the rocking or pitching of the ship, it is mounted on two sets of knife-edges at right angles to each other.

570. The Magnetic Elements. The detailed survey of the earth, to gain a more accurate and a more extended knowledge of its magnetic properties and the variation of them, has gradually become a very important work which is now carried on by nearly

all civilized governments. A complete statement of these properties at any location consists in specifying three distinct quantities, viz. the *dip*, or *inclination*, the *declination*, and the *intensity*. The first two determine the direction of the field, and the last quantity is its strength in dynes per unit pole strength.

571. The Dip. It has been assumed in what has gone before that the earth's field is horizontal. To test this, select a long thin unmagnetized steel bar and carefully poise it so that it swings freely in any plane and in any azimuth, but mounted with its center of gravity as nearly as possible at the point of support. Thus the needle will be in a position of nearly neutral equilibrium. Let the bar now be strongly magnetized and returned to its former place. It will then be found that the bar no longer hangs in an indifferent position, but that the positive pole is depressed in the northern hemisphere causing the axis to be inclined to the horizontal, the angle of *inclination* or *dip* as it is called being dependent on the latitude. In the neighborhood of Washington, D. C., at the present time, the dip is about 70°. Obviously the direction of the dipping needle at any place is the same as that of the earth's field.

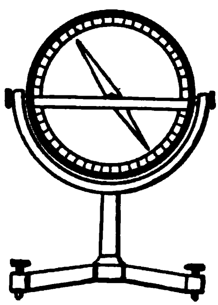


Fig. 367.

572. The Declination. Upon careful observation it is noted that the compass needle does not in general point due north, *i. e.*, that the vertical plane through its axis does not in general coincide with the plane of the geographical meridian through the place. The angle between these two planes is called the *declination* at the place. That at Washington is now about 4.5° west.

573. The Intensity. The intensity I of the earth's field at any point is measured like that of any other field, namely by the force in dynes acting on a unit pole. For many purposes it is useful to refer the total magnetic intensity to two components, a horizontal and a vertical one. Thus in Fig. 368 if Ot represent in magnitude and direction the total intensity, Oh and Ov, the sides of the rectangle of which Ot is the diagonal,

will represent the horizontal component H, and the vertical component V respectively, δ being the angle of dip. Obviously the following three relations exist,

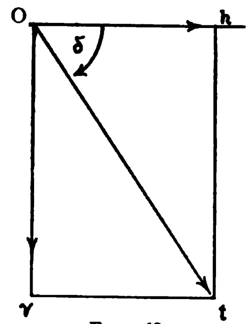


FIG. 368.

$$\frac{H}{I} = \cos\delta; \quad \frac{V}{I} = \sin\delta; \quad \frac{V}{H} = \tan\delta$$

I at Washington is now about 0.6 dynes per unit pole, and H is about 0.2 dynes per unit pole. H can be measured very accurately by the method given in § 564, and as δ can easily be observed by means of a dip circle (Fig. 367), V and I can be calculated. For the more refined methods of measuring the magnetic elements technical works on the subject should be consulted.

574. Magnetic Charts. If on a map a line be drawn through all points at which the dip has any fixed value, say 70°, and the same be done for other dip values, a system of lines will be developed analogous to the parallels of latitude. These lines are called *isoclinic* lines, or lines of equal inclination. Critical surveys of the dip show that as one goes further north the dip in general gets greater and that there is a point at which the dip is 90° and the needle is vertical. Such a point is called a magnetic pole of the earth.

This **north magnetic pole** has been from time to time located more or less exactly and is in the extreme northern part of North America. It was reached by Amundsen in 1907 and found to be at 75° 5′ N. latitude and 96° 47′ W. longitude. There is a corresponding south magnetic pole which the Shackelton expedition reached (1908) and found at latitude 72° 25′ S., and longitude 154° E. These poles are not stationary though their changes of position are very slow. The line through points of no dip is called the *magnetic equator.*

Lines of equal declination have the same general arrangement as the geographical meridians and are called *isogonic* lines. They evidently converge at the magnetic poles. The isogonic of zero declination is known as an *agonic line.* Fig. 369 shows the isogonic lines for the United States in 1900 as determined by the work of

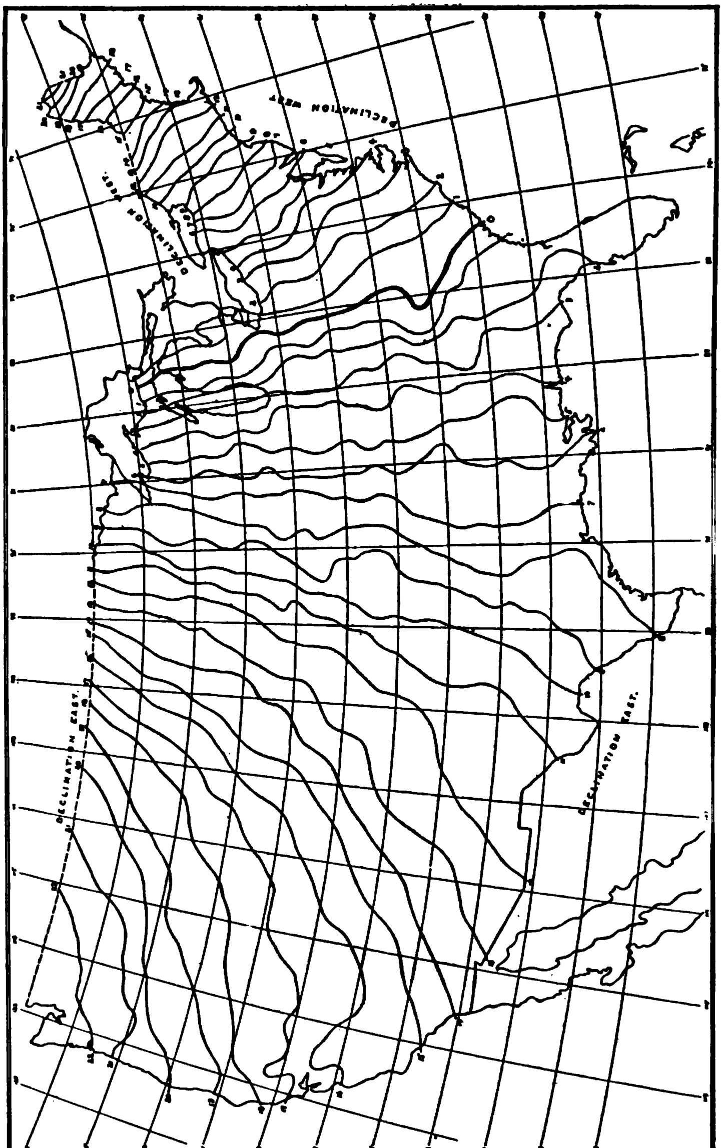


FIG. 369. Isogonic Lines for 1900.

the Coast and Geodetic Survey. Each of these lines is drawn from observations at stations many miles apart and hence give only the general course of the true curves. When the curves are

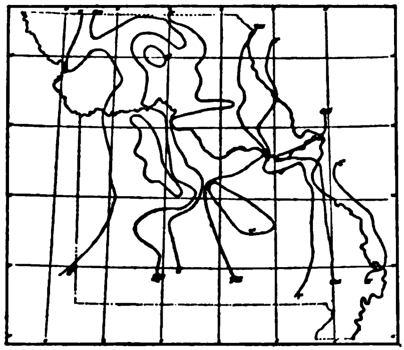


FIG. 370.

drawn from observations at stations close together they are found to be much more irregular. This is well shown by the chart (Fig. 370) for Missouri in 1890 drawn from work by Professor Francis E. Nipher of St. Louis. This, however, is a somewhat extreme case.

575. Variations of Magnetic Elements. Continuous records of these quantities at various stations have been kept for many years and these show that the magnetic elements at any one place are far from constant; that the law of variation is quite complicated; but that the changes are probably due to several different varying influences each of which is more or less periodic. In the case of a change of long period requiring many years to complete the cycle, the change is called the **secular variation.** Besides this there are two other periodic changes called respectively the **diurnal** and the **annual** variations. The former seems to depend on the earth's rotation and was discovered in declination by Graham in 1722, while the annual variation was observed by Cassini in 1870. Irregular variations of the magnetic elements sometimes take place and are called **magnetic storms.**

576. Continuous Automatic Records. The various magnetic elements, especially the declination, may be continuously recorded by attaching to the freely suspended magnet a light mirror from which is reflected a narrow beam of light. This is directed upon

a moving strip of photographic paper the motion of which is perpendicular to that of the magnet. The axis of time is thus along the paper and that of magnetic deflections is perpendicular to this. The result is an irregular curve, the ordinates of which measure the deflection of the magnet at any instant of time as located on the time axis. This recording instrument may be arranged for either the declination or for the dip. The photographic paper is usually rolled around a cylinder which is rotated uniformly by clockwork. Fig. 371 shows a record of an exceptionally violent magnetic storm at Kinguafiord in 1882.

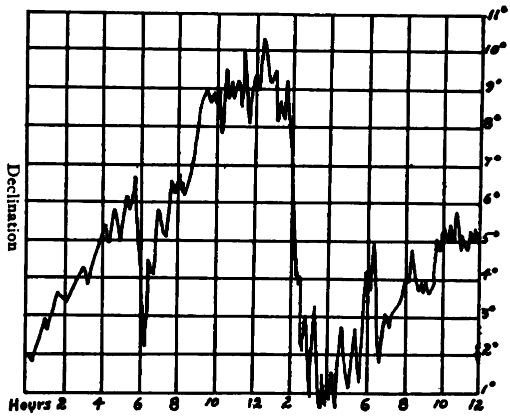


FIG. 371.

577. Theories of Terrestrial Magnetism. The earliest theory advanced to account for the earth's magnetism was that of Gilbert in his "De Magnete." He expressed his view briefly as follows: "Magnus magnes ipse est globus terrestris,"—the earth itself is a great magnet. Biot in 1805 suggested that the earth's magnetic phenomena could best be explained by supposing it to contain near its center a relatively short magnet. Gauss in 1833 attacked the problem mathematically and concluded that the earth contained a large number of irregularly arranged small magnets instead of one larger one. About 1850 Grover applied the result of the discovery by Oersted (§ 623) of electromagnetism, made in 1820, to explain the earth's magnetism. He assumed that it was caused by electrical currents circulating around the earth produced by the sun and influenced by the earth's motion, an explanation quite generally accepted at the present time.

ELECTROSTATICS.

Electrostatic Forces.

578. Electrification. It seems to have been known among the Greeks in the time of Thales (600 B. C.) that amber when rubbed acquires the property of attracting light bodies, such as bits of paper. The Greek word for amber is ἤλεκτρον, and so when Gilbert many centuries later (1600) studied the subject more in detail, he gave the name *electrification* to the property imparted by rubbing not only to amber but, as he found, to many other substances as well, such as sulphur, glass, resin, and the like. He thought that the metals and some other substances could not be electrified, and he called them *anelectrics*. It was later discovered by Gray, in 1729, that the true reason for this was because the metals are *conductors* of the electrification, and the latter was, therefore, led away to the earth and lost. All substances, therefore, may be divided into two classes, **conductors** and **non-conductors** or **insulators**, but we shall see that there is no sharp dividing line, the two groups merging into each other by imperceptible degrees if arranged in proper order. Nevertheless, as a class, certain materials like the metals are called good conductors, because they possess this quality in a marked degree, while other materials, as amber, sulphur, rubber, and glass, are called good insulators, because they conduct so very badly.

579. Electricity. Since the phenomena just described cannot be explained by reference to any other known property of matter, it has been generally agreed to assume the existence of something which is called *electricity*. On this hypothesis bodies which have been electrified are said to have acquired a *charge of electricity*, or more simply they are said to be *charged*. While it cannot be said at the present time that we know what electricity is, or its ultimate nature, or its relation to forms of gross matter, yet much has been done during the past decade towards solving all of these problems. This subject will be discussed more in detail in the sections devoted to "Radioactivity."

580. Eectrification of Two Kinds. This was noted by *DuFay* in 1733. If a stick of resin be rubbed with fur or flannel and brought near a light gilded pithball hung by a silk thread, the latter will be attracted. If it be allowed to come in contact with

the rod, repulsion will immediately take place. If now a glass rod be rubbed with silk and approached to the pithball, the latter will be attracted. In other words the glass and the resin act oppositely on the same pithball. They must, then, be in different electrical states, *i. e.*, there must be two kinds of electrification. The electrification of glass DuFay called **vitreous**, and that of the resin he called **resinous**. It was noticed, however, by Canton in 1753 that the rubbing body has as much to do with the result as the body rubbed. For example, if the glass be rubbed with fur or flannel, it will be resinously charged and will repel the pithball that it attracted when rubbed with silk. Again if the resin rod be rubbed with guncotton, it will be vitreously charged and will attract the pithball it repelled when rubbed with fur or flannel. In 1749 Franklin substituted the terms **positive** and **negative** respectively for vitreous and resinous, and in 1757 Wilke arranged the names of a number of substances in a series in such an order that each is positive when rubbed with any that follows. The following is a convenient series for reference.

1. Cat's fur.	5. Glass.	9. Wood.	13. Resin.
2. Flannel.	6. Cotton.	10. Metals.	14. Sulphur.
3. Ivory.	7. Silk.	11. Caoutchouc.	15. Gutta percha.
4. Quartz.	8. The hand.	12. Sealing wax.	16. Guncotton.

It is thus seen that most substances may be made either positive or negative at will by properly choosing the associated body; *i. e.*, glass is positive if rubbed with silk but negative if rubbed with fur.

581. DuFay's Law. Referring to the experiments on attraction and repulsion described above, it is to be noted that after coming in contact with the resin rod the pithball was repelled. If the same experiment be repeated with a glass rod rubbed with silk and another pithball, the latter will also be repelled by the glass. If now the two balls be brought near each other, they will strongly attract each other. Moreover the glass rod will attract the ball that is repelled by the resin and *vice versa*. These phenomena are consistently classified by supposing that, at first, the uncharged ball, coming in contact with the rod attracting it, acquires a charge like that on the rod, and that subsequent repulsion is due to the fact that *like kinds of electrification cause mutual repulsion* of the

bodies charged. Again, since the two attracting pithballs must be oppositely charged, we are led to the conclusion that *unlike electrifications cause mutual attraction*. This generalization was first announced by DuFay.

582. Mechanism of the Process. In all cases of electrification thus far considered, the act has been described as one of rubbing, and, from this idea that rubbing or friction was an essential feature of the process, the term *frictional electricity* was for many years used. As a matter of fact, however, friction is only an accident. The effect depends on the contact and subsequent separation of the bodies composed of two different substances, or, as is sometimes the case, even of the same substance in different physical states. Thus if a strip of white ribbon and one of black ribbon be drawn together between the dry fingers, the white piece will be positively and the black negatively charged. As more points can be brought into contact and separated by rubbing than by any other simple act, this came to be the method employed.

If the bodies used be non-conductors the charge stays where it is developed and is even accumulated by multiple rubbings. If however they be conductors, the two opposite charges flow together at points which are last to separate and the bodies are thus discharged. For this reason it was at first thought that only such substances as glass, resin, rubber, etc., could be charged. Since, however, a long metal wire or other conductor if insulated may be charged at one end and exhibit the usual phenomena of electrification at the distant end no matter how remote, and since investigation has shown that *the act of transferring* the charge along the conductor is attended by interesting and important phenomena, both within and without the conductor, the subject of electricity is divided into two parts, *electrostatics,* dealing with charges at rest or in equilibrium, and *electro-kinetics* dealing with charges in motion. For the present we shall concern ourselves with electrostatics only.

583. Coulomb's Law. Coulomb in 1784 first investigated the law of force between electric charges. He demonstrated, with considerable accuracy, that the force exerted between charged bodies at a distance is directly proportional to the product of the amounts of charge and inversely proportional to the square of

the distance between them. For the charged bodies very small conducting spheres are necessary, in order that the size of the bodies may be small as compared with the distance between them. In this case the distribution of the charge on either body will not be appreciably affected by the charge on the other. If two such small spheres be charged with quantities q and q' of electricity and placed at a distance, r, apart, the force exerted between them in air under usual conditions will be expressed by

$$F = \frac{qq'}{r^2}$$

with the same agreement as to signs as in § 558. Since the actual force depends on the separating medium, a constant, known as the *dielectric constant,* must be inserted. This does not vary for any one medium and is independent of the charge and of the distance. The general expression then becomes

$$F = \frac{qq'}{kr^2}$$

k having a value of unity for air. From these relations the unit charge may easily be defined. It is that quantity which would exert upon an equal quantity a centimeter distant in air a force of one dyne. This is known as the **C. G. S. electrostatic unit quantity of electricity.**

584. Energy Relations. Since the forces of attraction and repulsion resulting from electric charges cause masses to start from rest and acquire considerable speed, or even to raise themselves in opposition to their weights, electrification implies energy. This energy, assuming the principle of the conservation of energy (§ 318), must be acquired by the bodies in the process of electrification. Further investigation has shown that this energy is conferred upon the system, composed of the two bodies rubbed and the medium between them, in the act of separating the bodies oppositely charged and therefore mutually attractive. It is much as if the bodies were joined by fine rubber threads, which would require work to be done in stretching them, and which would tend to draw the bodies together again if separated. The medium between the bodies is, therefore, the seat of the energy and we define an *electric field* as the space surrounding a charged body, just as in the case of the magnetic field about a magnet.

585. Electric Field. The intensity at any point of an electric field is defined as the force exerted by the field on a body carrying a unit charge and placed at that point. As in the case of a magnetic field, we represent an electric field by **lines of force** so distributed, that the intensity of the field is *proportional* to the number of lines per square centimeter of a surface perpendicular to the field. (In the case of an *electric* field the lines are, for certain reasons, so chosen that the number per cm^2. equals the intensity divided by 4π.) The positive direction of a line of force is that in which a small body would tend to move, if positively charged and placed there. It is clear that a negative charge would tend to move in the opposite direction. When a field is caused by two opposite and separated charges the lines of force start at points of the positive charge and end at corresponding points on the negative charge. The arrangements of the lines in electric fields are shown in several specific cases by the accompanying diagrams. Fig. 372 represents the field generated by a single positive charge concentrated on a very small body. The lines of force are straight radial lines emanating from the charge. Fig. 373 shows the arrangement when two equal unlike charges are located on two near bodies; while Fig. 374 shows the result of the action of two equal like charges. No two lines can cross for then a charge placed at the point of meeting would tend to move in two directions at the same time. Examination of the figures will suggest what Faraday first called attention to, viz., that the intervening medium plays a very important part in the phenomena of elec-

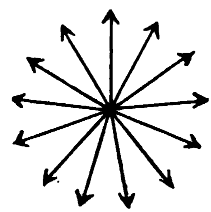


Fig. 372.

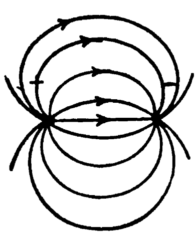


Fig. 373.

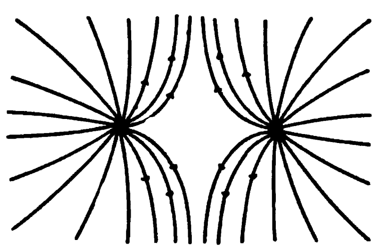


Fig. 374.

trification. To this medium he gave the name, *dielectric*. In an electric field Faraday pictured the lines of force as tending to become shorter like stretched elastic threads and also mutually repellent in a direction perpendicular to their lengths. Electricity having been defined as the agent to which is due all the phenomena of electrification, it may be said that the latter is to be viewed as a state of dielectric strain brought about by the forcible separation of two bodies oppositely charged.

586. Electric Convection. If a light conducting ball suspended by a silk thread be brought between two oppositely charged surfaces, Fig. 375, it will be attracted towards the one it happens for the moment to be nearer. On coming in contact with that surface, it acquires a part of the latter's charge by conduction and is therefore immediately repelled. Since now it is also attracted by the charge on the other surface, it will quickly come in contact with it, have its own charge reversed and be again repelled. The little ball will thus vibrate back and forth between the two charges, gradually discharging both bodies. This kind of actual transfer of electricity by the motion of a charged carrier is called *electrical convection*. Franklin arranged a set of bells in pairs, one of each pair being connected with his lightning rod and the other to the ground. Between the bells of each pair a vibrating metal ball was suspended. On the approach of a thunder storm the bells would become charged and ring, owing to the action above described.

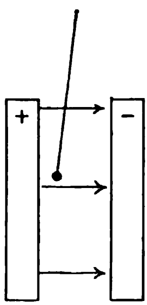


FIG. 375.

ELECTRICAL POTENTIAL.

587. Potential Analogies. Analogous relations in the various departments of physics are extremely useful and in the study of electrical *potential* we can derive much assistance by comparing it with the analogous quantities in hydrostatics, in aërostatics, and in heat. Liquids for example naturally flow from places of higher to those of lower *level* and difference of level in hydrostatics is analogous to difference of potential in electrostatics. Again if a gas be pumped into a receptacle the *pressure* increases continuously up to the point when the vessel either bursts or leaks. So, when

a conductor is charged with a larger and larger quantity of electricity, the *potential* rises continuously till the strength of the surrounding dielectric medium is overtaxed and a leak takes place consisting of a *brush discharge,* as it is called, or perhaps the dielectric gives way all of a sudden at some point and a *disruptive discharge* or spark takes place. In either case a part of the charge escapes to a place of lower potential, and the potential of the conductor is lowered to the point at which the dielectric is strong enough to hold the charge on the conductor. Similarly heat flows from places at higher *temperature* to those at lower temperature, and here temperature is the analogue of electric potential.

Further, just as the sea-level is arbitrarily taken in practice as the zero of gravitation level for convenience of reference, so the earth as a whole, being an enormous conductor, is conveniently referred to as being at zero electrical potential. Any conductor, therefore, joined to the earth by a wire or other conducting means, has its potential reduced to zero.

588. Electrical Potential. The magnitude of the electrical potential at any point in space is defined as being numerically equal to the work necessary to move a unit quantity of positive electricity from infinity to the point in question against the forces of all the electricity in existence. A unit charge, then, placed at such a point would possess potential energy, which would tend to diminish by the moving of the charge to another point at which the potential is of less value. In other words positive charges tend to move from points of higher to those of lower potential, and the resultant force at any point is in the direction in which the potential diminishes most rapidly. From the definition of potential, it follows that the difference of potential between two points is numerically equal to the work necessary to convey unit positive charge from the point at the lower potential to that at the higher. Therefore a C. G. S. **unit potential difference** exists between two points, when an erg of work is required to carry a unit positive charge from the lower to the higher potential. From this we reason at once that if V represent the constant potential difference between two points, and if Q be the charge conveyed, the work done will be expressed by the equation $W=QV$. This assumes that the movement of Q does not appreciably change the rest of the electrical distribution.

589. Equipotential Surfaces. Suppose a unit positive charge concentrated at a point, *A*, Fig. 376, and isolated from the influence of all other electricity. Lines of force emanate from it radially in all directions. About *A* in every direction there are evidently points at such a distance as to require an erg of work to move a unit positive charge from each up to *A* against the repulsion of the charge placed there. These points, being all at the same distance from *A*, must lie on the surface of a sphere with *A* as center. In other words the potential difference between *A* and every point of the spherical surface described will be unity. Again, a second spherical surface outside the first may be drawn all over which the potential will be two units different from that at *A*. Such surfaces as these are called *equipotential surfaces*. In the simple case supposed for illustration the system consists of a series of concentric spherical surfaces, at such a distance apart as to require an erg of work to be done on a unit positive charge, in order to convey it from any one to the next. Such a series of spheres is shown in the figure.

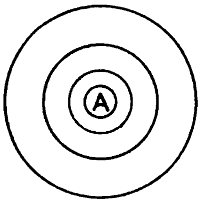


FIG. 376.

In general the actual distributions of electricity to which the field is due are not nearly so simple, and the systems of equipotential surfaces resulting may be quite complicated. Since at all points of an equipotential surface the numerical value of the potential is the same, no work is required to move a charge along the surface; but in a field of force the only direction in which a charge can be moved without work is perpendicular to the lines of force. These, therefore, at every point must be perpendicular to the equipotential surfaces. No two such surfaces of the same system can intersect; if they could, at the points common to both we should have two different potential values at the same time—an impossibility. Again, in all cases the work required to move a charge from *any* point of one equipotential surface to *any* point of another is quite independent of the path along which the motion takes place. This is obvious if one considers that in passing by an oblique path from one surface to another, while the distance is longer the force is less in the same ratio; therefore the product of the two, or the work done, is constant.

590. Intensity as a Function of Potential. If V and $V + \Delta V$ be the potentials at two very near points, A and B, on a line of force, Fig. 377, and if Δr be the distance between these points, we can express in two ways the work done on unit charge in moving it from A to B. If $\bar{f}$ be the average intensity at the points situated between A and B, $\bar{f}\Delta r$ will be the work done in moving unit charge from A to B against the force. But, from the discussion of potential difference, ΔV also represents the same amount of work. We have, then, since the direction of the force is the same as that in which V *decreases*, $\bar{f}\Delta r = -\Delta V$, or $\bar{f} = -\Delta V/\Delta r$. If A and B be made to approach each other till they coincide, the left-hand side of the above equation, $\bar{f}$, will be the actual value of the intensity at the place of meeting, and the right-hand side, $-\Delta V/\Delta r$, will approach the value $-dV/dr$. In other words this means that the resultant intensity at any point is numerically equal to minus the rate at which the potential changes along the line of force through the point.

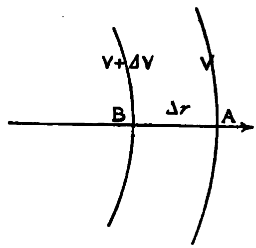


FIG. 377.

591. The Potential of a Conductor. In all that has been said heretofore, potential has been discussed as a condition of things *at* a point, and this is the only strictly proper way to regard it. If a charge be given to an insulated conductor at one or more points, the charge at once will arrange itself in a definite way and, in all its parts, will assume a state of stable equilibrium. This follows immediately from the preceding discussion. There can be no resultant force other than that perpendicular to the surface; otherwise the tangential component would move part of the charge at some point, and this would continue till the stable arrangement spoken of has come about. Since, then, the surface is everywhere perpendicular to the force, it must be an equipotential surface. In other words, the surface of a charged conductor in equilibrium is always equipotential, and it is often convenient to ascribe the value of the potential at each of its points to the conductor as a whole, and to speak of the conductor as having been *charged to a certain potential*, or even as *having* this potential. If a conductor then be charged to any potential V, this means that it has on its surface a charge $+Q$, of such a value as to require V ergs to be expended in bringing a unit positive charge to *any* point of the surface from infinity against the repulsion of Q. If a higher

potential is to be obtained, Q must be increased, and in general Q is proportional to the resulting potential which is due to it, *i. e.,*

$$Q = CV,$$

in which C is a constant depending on the size, shape, or environment of the conductor. From the formula it appears that C is numerically equal to the charge required to change the potential of the conductor by unity, or to raise its potential from zero to one. This quantity, C, is called the **capacity** of the conductor. C also depends on the dielectric constant, k, of the medium in which the conductor is immersed; in fact it varies directly as this constant. For if Q be kept constant and k increased in any proportion, the repulsion of Q on the approaching unit charge will be decreased in the same proportion (§ 583). Hence the potential will be decreased, and therefore the capacity increased, in the proportion in which k is increased.

592. Potential at Points Within a Charged Conductor. Since the surface of a charged conductor is equipotential, all points within it must be at the same potential unless there be another charge there. For otherwise there would be points at higher or lower potential than that at the surface; but lines of force must start from positive charges of which there are none by hypothesis. Similar reasoning shows that there are no points at lower potential than that at the points of the surface. The potential is therefore constant throughout the space within a charged conductor.

593. Force at Points Within a Changed Conductor. Since the intensity at any point is equal to the rate of potential change, *i. e.,* to $-dV/dr$, it follows that wherever the potential is constant the force is zero and hence there is no force within a charged conductor. It can be further demonstrated that this condition can only be fulfilled if Coulomb's law is true. The simplest case to consider is that of a uniformly charged sphere, Fig. 378. Let us find the resultant intensity at point P within it, chosen at random. Denote the charge on each unit of surface by σ. Let P be the vertex of a double cone of small angle intercepting the charges $s_1\sigma$ and $s_2\sigma$ at the respective distances of r_1 and r_2. The intensity at P in one direction is $s_1\sigma/r_1^2$ and in the other $s_2\sigma/r_2^2$. But from geometrical considerations $s_1/r_1^2 = s_2/r_2^2$. As regards, then, the effect of the charges intercepted by the small cone chosen, the resulting force is zero. But the whole interior of the conductor may be regarded as made up of such cones. Therefore the force at all points within the surface

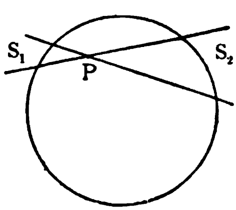


FIG. 378.

is zero, on the assumption of Coulomb's law, or the law of inverse squares. Since experiment shows that at all such points as P the force *is* zero, the law of force from which theoretically this would follow must be the *true law*.

594. Value of Potential Due to Charge at a Point. Let a charge Q, Fig. 379, be concentrated at a point A; let us calculate the work a unit charge would accomplish in moving under the repulsion due to Q from P_1 to P_2. Suppose that at any instant it is at P, at a distance r from A. In

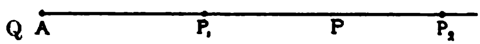


FIG. 379.

moving a distance Δr, the work done (assuming the dielectric to be air) is $Q/r^2 \cdot \Delta r$, and therefore in moving from P_1 to P_2, the work must be

$$Q\int_{r_1}^{r_2}\frac{dr}{r^2} = Q\left(\frac{1}{r_1}-\frac{1}{r_2}\right)$$

If P_2 be at an infinite distance, $1/r_2$ will be zero and the expression for work becomes Q/r_1. Hence this is the work required to bring up to P_1 from infinity a unit positive charge against the repulsion of Q, and thus is by definition the potential at P_1 due to Q at a distance r_1 from P_1. From this it follows that, if Q be composed of distributed parts, q_1, q_2, q_3, etc., at distances r_1, r_2, r_3, etc., respectively from any point P, the potential at P due to the total distribution will be

$$\frac{q_1}{r_1}+\frac{q_2}{r_2}+\frac{q_3}{r_3}+\text{etc.}$$

or briefly $\Sigma q/r$, which is expressed more exactly by the integral $\int dq/r$, in which the limits must take in all the charge concerned wherever it may be or in whatever way distributed. If the medium be other than air and has a dielectric constant k, then the expression becomes

$$V = \frac{1}{k}\int\frac{dq}{r}$$

595. Capacity of a Spherical Conductor. As a special case of the above, consider a uniformly charged sphere and locate the point P at its center. The potential at P due to the charge will be given by the equation

$$V = \int\frac{dq}{r}$$

Since r is constant and equal to the radius of the sphere, a, the above expression becomes

$$\frac{1}{a}\int dq = \frac{Q}{a}$$

Therefore $Q = aV$ expresses the relation between the charge and the potential at any point within the conductor, since the potential at all such points is the same. But in general $Q = CV$, in which C is the capacity of the conductor (§ 591); therefore in the special case of a spherical conductor, $C = a$ and **its capacity in air is numerically equal to its radius.**

596. Potential Energy of a Charge. Although strictly the energy of a static charge is distributed in the surrounding dielectric medium, it is possible to express its value in terms of the charge, Q, and of the potential, V, of the conductor carrying the charge. We have seen (§ 588) that the work required to take a charge Q from one point to another against a potential difference V is QV. In charging a conductor, however, the charge may be thought of as being added a little at a time and as increasing from zero gradually to its final value Q. The potential in consequence must increase from zero uniformly to its final value V, and the average potential, then, through which Q has been raised is $\frac{1}{2}V$. The work of charging the conductor is, therefore, $\frac{1}{2}QV$, and this represents the energy of the charge Q. Since $Q = CV$, we may express the energy as follows,

$$E = \tfrac{1}{2}QV = \tfrac{1}{2}CV^2 = \tfrac{1}{2}(Q^2/C)$$

ELECTROSTATIC INDUCTION.

597. Mechanism of Induction. If a charged body, A, Fig. 380, is brought near an uncharged insulated conductor, BC, the latter is found to be electrified. The side or parts nearest to the charged body are electrified oppositely to the body while the more remote parts are similarly charged. Electri-

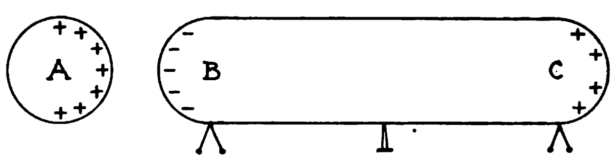


Fig. 380.

fication brought about in this way without any contact is called *induction*. One explanation consists in supposing that the neutral body naturally possesses equal amounts of positive and negative electricity which are separated by the attraction and re-

pulsion of the charged body when the latter is brought near, the unlike kind being attracted to the nearer side while the like kind is repelled to the remote parts. This is consistent with the following fact, viz. that if the charged body be now removed, the conductor loses entirely its electrification, this resulting from the recombining of the two kinds of charge.

Another way of explaining the above facts is the following. Suppose a positively charged body at *A*, Fig. 380, in virtue of which all neighboring points, as, for example, *B* and *C*, are at definite potentials depending on their distance from *A*. *B*, being near, is at a higher potential than *C* which is farther away. Suppose now an uncharged conductor be brought and placed so that *B* becomes one point of the surface at one end and *C* a point at the other end. Since *B* and *C* are now points on a conductor their potentials must be equal. Hence there must be a negative charge around *B* that lowers its potential and a positive one at *C* that raises its potential. Again, since originally the potential at *B* was higher than at *C*, the direction of the electric force was from *B* to *C* and, therefore, when the conductor made it possible, a positive charge moved in the direction *BC*, while a negative one moved in the opposite direction.

598. Effect of a Conductor on a Field. Fig. 381 shows in a general way the effect of bringing a conducting sphere *A*, initially uncharged, into the radial field produced by a positively charged sphere *B*. The potential at all points of *A* is that which was formerly the value at the point now occupied by its center. As the field intensity within the conducting surface of *A* is everywhere zero, no lines of force cross the boundary, but the same number leave on the side away from *B* as abut on the side near *B*. In all cases they approach and leave *A* in a direction perpendicular to its surface since *A* is an equipotential surface. The equipotential surfaces are shown everywhere per-

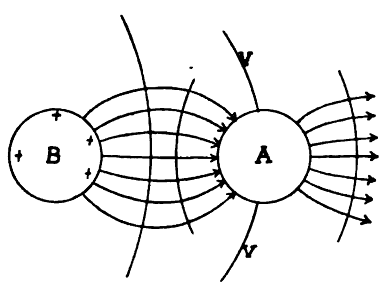


FIG. 381.

pendicular to the lines of force, the surface of *A* becoming now a part of the one, *VV*, that formerly passed through the center of *A*. It follows from the reasoning of § 597 that a negative charge has moved to the left of *A* and a positive charge to its right.

599. Changes Produced in a Field by an Insulator. If, instead of the conducting body *A*, an insulator of greater dielectric constant than the surrounding medium be brought into an electric field, the distortion of the lines of force will be something as shown in Fig. 382. The effect is much the same as that produced on a magnetic field by bringing a bar of iron into it. The lines of force are more crowded through the non-conductor than formerly through the space now occupied by it.

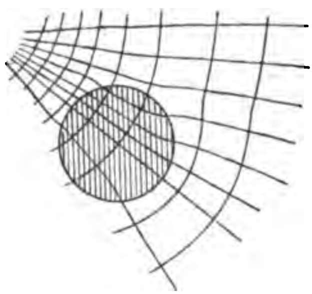


FIG. 382.

600. Zero Potential of the Earth. On account of the enormous electrical capacity of the earth and its general good conductivity, its potential cannot differ greatly from time to time or from place to place. For this reason it has been agreed for convenience to calculate potentials from that of the earth assumed as zero. Since any two conductors joined together by a conducting medium, as a wire for example, are quickly brought to the same potential, any conductor joined to the earth has its potential brought to zero. The usual practical method of "earthing" a body is to join it by wire to the nearest gas or water pipe. In the absence of both of these the wire must be led to a large metal plate or grid buried in the ground.

601. Charging by Induction. The method of charging a body, referred to in § 578, was by conduction. The result was as if the new body had acquired some of the electricity that had just before resided on the charging body. The new body may however be charged and yet not be brought into contact with the charging body at all. Referring again to Fig. 380 it is seen that the right end of *BC* is positively charged while the left end is negative, because of the inductive action of *A*. If, now, *BC* be touched with

the finger or in any other way brought into conductive communication with the earth, the repelled positive electricity will quickly take this path of escape to the earth. If now the earth connection be broken, BC will be found to be *negatively* charged even if A is removed from the vicinity. This is known as *charging by induction,* and the character of the charge is *opposite* to that of the charging body. In view of this inductive action even the phenomena of conduction may receive a different explanation from the one already advanced. Suppose A be brought up into contact with BC at B. The two opposite and mutually attractive charges will combine, a small spark passing just prior to actual contact. If, now, the bodies be separated, BC will be positively charged, but the charge is the positive charge that was originally associated with the negative charge that has just combined with the positive of A on contact. In other words, the positive charge now on BC is not a part of that which was formerly on A and which has passed over to BC; it was always on BC but has been made evident or *free* by taking away from it the same amount of negative charge with which it was formerly associated. Thus we see that it is logical to assume that conduction is preceded by induction. Induction thus plays a very important part in the phenomena of this branch at least of electrodynamics.

The charging by induction may also be explained by considering it as an action of the intervening medium. Lines of force pass out in all directions from A, Fig. 380. When the conductor BC is brought into the field lines abut at the left hand side of BC and an equal number leave at the right hand side, *i. e.,* the former side is charged negatively, the latter positively. Upon touching BC connection by a conductor is made with the earth and the stress of the medium, represented by the lines from BC to the earth, is removed. After breaking the connection with the earth there remain only the lines passing from A to BC and the latter's charge will therefore be negative only.

602. The Gold-leaf Electroscope. This instrument (Fig. 383), first constructed by Bennet in 1787, consists of a pair of gold-leaf strips about 5 mm. wide and 5 cm. long, each with one end attached to a metal rod 10 cm. or more long, terminating at the top in a plate or ball. These leaves are enclosed for protection in

a suitable container, often a glass jar, through the insulating top of which the rod passes and is fixed, the plate or ball being on the outside for receiving the charge. As a still further protection against local charges on the glass container and inductive influences from outside a light wire cage is sometimes placed within the glass jar and connected with a small terminal on the outside. By attraction this cage also helps to diverge the leaves. If, from misjudgment or accident, the leaves are given too heavy a charge, they will spread out and come in contact with the cage and will thereby be discharged and will fall back in place. As a divergence of the leaves indicates a difference of potential between them and the cage, if the latter be made zero by connecting to earth, the amount of divergence will, in a general way, measure the potential of the leaves or the body from which they were charged.

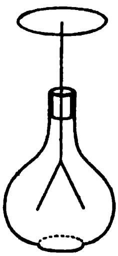


FIG. 383.

If a stick of sealing wax, charged by rubbing it with flannel, be brought toward the plate or ball at the top, the leaves will diverge with a negative charge by induction. If the plate be connected to the ground for an instant, by touching it with the finger, for example, the leaves will collapse, the negative charge being driven to earth by the repulsion of the charge on the stick of wax. If now the latter be removed after the ground connection has been broken the leaves will diverge again, but this time with a positive charge. If a positive charge be brought near the plate of a positively charged electroscope, the leaves will diverge more, the charge of the instrument being repelled to them. If, however, a negative charge be brought near, the leaves will collapse because their positive charge will be attracted upward to the plate by the action of the opposite charge in the vicinity. In general, then, which ever way the electroscope is charged, a similar charge will cause an increased effect, while an opposite charge will cause a diminished effect. An electroscope is thus capable of indicating the character of a charge as well as its presence.

603. Distribution of Charge. If a body be a non-conductor, any charge of electricity communicated to it must necessarily stay where it has been placed. If it be a conductor properly insulated, the charge will distribute itself in a very definite way,

depending on the shape of the body and upon the location of other charges, if any, in the neighborhood. In any case the charge will be located on the surface of the body, never within its mass. If, for example, the body be a charged conducting sphere and if there be no disturbing charges near it, the charge will be uniformly distributed all over the surface. The uniformity of the distribution might be expected from the symmetry of the spherical form. The superficial character of the distribution is not so obvious, however, though of much importance. The fact can be demonstrated experimentally in several ways. Franklin, for example, electrified a small silver can on an insulating stand and lowered within it a cork ball by means of a silk thread. The ball was not attracted, though it had been when brought near the outer surface of the can. Even when the ball was made to touch the inside surface of the can, it was uncharged when withdrawn. If, however, it be made to touch the outside, it acquires a charge. Franklin himself did not understand the reason for this, but it may be explained on the assumption that all parts of the charge are mutually repellent and, therefore, it will arrange itself in such a way as, on the whole, to be most widely distributed. Coulomb also, by means of his torsion balance, Fig. 356, explored the surface of an electrified cylinder in which cavities had been made. He used a little disk of gilt paper mounted on a shellac rod for a handle, in order to convey the charges from different points in the cylinder to the torsion balance. This device is called a *proof plane*. In no case was he able to find any evidence of a charge within the cavities. Faraday's method of testing this question was on a much larger scale. He built an insulated cubical box, about 8 feet on an edge, and covered it with a conducting layer of tinfoil and copper. It was charged by a powerful electrical machine (§ 618) until long sparks and brushes escaped from its corners and edges. While it was in this state Faraday himself stayed within the cube and used various tests for the presence of an electrical charge, but was unable to detect the least charge on the inside of the cube.

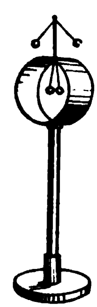


FIG. 384.

A simple and effective lecture table device for showing the same thing consists of a hollow metal cylinder, *A*, Fig. 384, mounted on an insulating

stand B. A pair of pithballs P are hung by conducting threads from a point on the inside of the cylinder and a similar pair P' on the outside. On charging the cylinder it will at once be noticed that the balls P' will mutually repel and will stand out at an angle depending on the amount of charge given the cylinder, but that balls P will be quite unaffected, thus indicating that the charge is distributed entirely on the outside of the conductor.

604. Faraday's Ice-pail Experiment. This experiment was first made by Faraday and is known as "Faraday's ice-pail experiment" (because he actually used a pewter ice-pail). Place an insulated metallic vessel in communication with an electroscope as in Fig. 385. Into this lower a charged body by means of a silk thread, or other non-conductor. The leaves will diverge but the divergence is *not changed* by moving the charged body about within the cavity, *or even by bringing it in contact with the side of the vessel.* If the body be then withdrawn it will be found to be uncharged. As the divergence of the leaves was not affected, no change in the charge on the outside of the vessel was caused by bringing the charged body in touch with it. This shows that when the charged body is lowered into the vessel, there is produced on the interior surface of the conducting vessel by induction a charge of *equal* value with, and of *contrary* sign to the charge of the body introduced. The actual distribution of this induced charge depends on the form and position of the inducing body. There is also induced on the outside of the vessel a charge of *equal* value and of *similar* sign to that of the inducing body. Hence when the inducing body is made to touch the interior of the vessel two equal and opposite charges combine and neutralize each other. The body when withdrawn is then found to be uncharged and the induced exterior charge is of course unchanged by the combination of two equal and opposite charges on the inside. This and other evidence indicates that induced charges always occur as pairs of equal and opposite charges and that

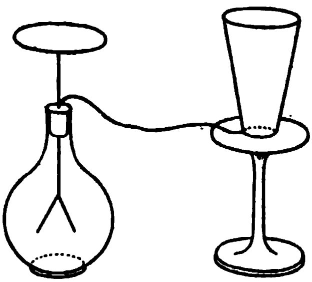


FIG. 385.

when one body gives a charge to another by conduction, the total charge of the two is unchanged.

This can also be explained by the changes produced in the medium. The lines originating from the charged body end on the surface of the earth or on bodies not insulated from it. If the body is now lowered into the insulated vessel we may distinguish two independent fields under electrical stress, the one inside, the other outside the vessel. Lines pass inside from the charged body to the vessel; and the same number outside from the vessel to the earth. Moving the charge inside the vessel cannot produce any change whatever outside. As soon as the charged body touches the wall of the vessel, the stress inside disappears and no charge is left, while on the outside there remains the same charge as before, equal to that of the originally charged body.

605. Conservation of Electricity. If two bodies in their natural state, such as a piece of silk and a piece of glass, be lowered into a hollow vessel like a can placed on the plate of an electroscope, no indication of a charge will appear; *i. e.,* the leaves will not move even though the glass and silk be rubbed together and separated as if to charge them. If, however, either be removed, the leaves will diverge because of the presence of a charge on the body left within the vessel. Since, when both bodies were within the vessel, no charge was indicated even though a test of each separately showed them both to be charged, it follows conclusively that the bodies must have been charged *oppositely* by exactly equal amounts. That is, when a charge of either kind is developed anywhere by contact, an opposite charge of just the same amount must also be developed. From this and the preceding paragraph we conclude that *the total amount of electricity of both kinds is always zero.*

606. Surface Density. In the case of an electrified body the charge on its surface may be tested at various points by means of the electroscope. To do this a small conducting sphere or disc, attached to an insulating handle, is brought successively in contact with different points of the body and, after each contact, touched to the electroscope. It will be found, in general, that the deflections produced are not the same. If the body be a conducting sphere the test shows that the distribution is uniform all

over the surface. If however the shape be that of an ellipsoid or ovoid an examination of the surface will reveal the fact that at places of sharper curvature the charge is greater, and hence in the figures mentioned the charge will be much greater at the ends than at the flatter sides. The quantity of electricity residing on a square centimeter of a surface is called the *surface density* of the charge and is denoted by σ. In the case of a spherical distribution it is clear that the surface density is the same at all points, while in most other cases this is not true, and then it is defined as the limit of the ratio of the charge on a small area about the point in question to that area as the latter is made smaller and smaller.

From the above it follows that at actual points the electric density must be very great, and this is a well known observed fact. Since, moreover, the charge on a point is repelled by the charge on all other parts of the conductor it often happens, and will always happen if the point be sharp enough, that the stress in the dielectric is sufficient to break it down and the electricity escapes from the point. (See Radioactivity.) Franklin was the first to recognize the value of pointed conductors in discharging electricity harmlessly into the air and suggested their use at the ends of lightning rods in place of the balls which had been formerly used.

607. Effect of Induction on Capacity of Condensers. The capacity of a single conductor uninfluenced by others has already been defined as the ratio of its charge to its potential (§ 591), or in other words as equal numerically to the charge necessary to raise its potential from zero to unit value. In the case of a sphere the capacity in C. G. S. units is equal to its radius in centimeters. If, now, such a sphere be enclosed by a concentric conducting spherical shell of slightly greater radius, the two being separated by a dielectric and the shell being connected to earth, the capacity will be found to be very greatly increased. The capacity of a conductor is, therefore, a function not only of its size and shape but also of its environment. Such a combination as is described above is called an *electrical condenser*. Hence an electrical condenser is a device for increasing the charge on a conductor without increasing its potential.

608. Calculations of Capacities. Let the above sphere, Fig. 386, have a radius of a centimeters and let the enclosing shell have an inner radius of b centimeters with a small air space between. If a charge q be given to the sphere, a charge $-q$ will be induced on the inner surface of the shell. Since it can be proved that charges distributed uniformly over spherical surfaces affect outside points as if the charges were concentrated at the center, O, of the surfaces, it follows that the potential at P at a distance r from O, due to q and $-q$, is

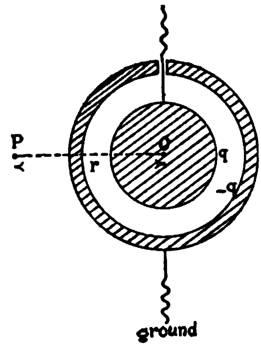


FIG. 386.

$$\frac{q}{r} - \frac{q}{r}$$

This is evidently zero. If P be taken between the sphere and shell, the potential due to q is q/r as before; but that due to $-q$ is $-q/b$, since this last is constant throughout the space enclosed by the shell and equal to the value it has on the surface (§ 592). The total potential is therefore equal to $q/r - q/b$. At the surface of the sphere

$$V = q\left(\frac{1}{a} - \frac{1}{b}\right)$$

since in this case $r = a$. At all points within the sphere the potential is constant and has of course the same value as at the surface. Since the capacity is equal to the charge producing unit potential, we see at once from the above formula that

$$C = \frac{1}{\frac{1}{a} - \frac{1}{b}} = \frac{ab}{b - a}$$

As $b - a$ is evidently the thickness of the dielectric, which we will call d, and as this is supposed small, ab will differ little from a^2 so that $C = a^2/d = 4\pi a^2/4\pi d = S/4\pi d$, where S is the surface of the sphere.

From the above we can also derive an expression for the ca-

pacity of a condenser consisting of two parallel plates. For they may be regarded as parts of two very large concentric spheres forming a spherical condenser. Hence the capacity of such a plate condenser is also expressed by the formula $C = S/4\pi d$, where S is the area of each plate and d their distance apart (supposed small). If the space between the conductors is not air but a material of dielectric constant k, then it follows from § 591 that $C = kS/4\pi d$.

It is at once evident that the capacity of either a spherical or a plate condenser may be increased indefinitely by decreasing d. The limit is reached in any practical case when the thickness is not sufficient to withstand the potential required and a spark breaks down the dielectric. By making a simple calculation, it will be seen that a 10 cm. sphere has its capacity increased more than 100 fold by surrounding it with a grounded shell 1 mm. away, the space between being air. In practice condensers are very frequently composed of metallic sheets, such as tinfoil, separated by large sheets of some dielectric, such as glass, mica, paraffined paper, or oil. Fig. 387 shows the arrangement. The sheets of odd numbers are joined together and earthed, while the even ones are joined together and receive the charge. One set is completely insulated from the other by layers of the chosen dielectric, especially selected with reference to its resistance to breaking down under high potential.

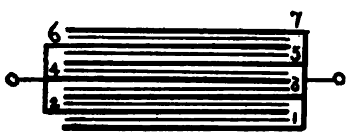


Fig. 387.

609. Field Between the Plates of a Condenser. If the distance, d, between the plates of a parallel-plate condenser is small compared with the dimensions of the plate, the lines of force between the plates are parallel, and the field is, therefore, uniform. If the distance between the plates be decreased (the charges remaining unchanged), the only change will be a shortening of the lines of force, without any change in their number. Hence the intensity, f, of the field will remain unchanged and so also will the force of attraction, F, between the plates. The difference of potential, V, is the work done in taking a unit charge from one plate to the other. Hence $V = fd$ and $f = V/d$. From this other expressions for f can be derived. For, if C is the capacity of the condenser and q and $-q$ the charges on each plate $q = CV$ and $V = q/C$. Hence $f = 4\pi q/S = 4\pi\sigma$ where σ is the surface density.

610. The Leyden Jar. The principle of the condenser is said to have been accidentally discovered by von Kleist in 1746. While electrifying some water in a bottle held in the hand, he received a shock on taking hold of the nail placed inside to make contact with the generator. The hand had formed one side of the condenser and the water the other, while the glass of the bottle formed the dielectric. When he touched the chain in the water with the other hand, the discharge took place through his body and produced the shock. A more practical form of this sort of condenser resulted in the so-called Leyden jar, Fig. 388. It consists of a glass jar coated about two thirds the way up outside and inside with tin foil. Through an insulating top passes a metal rod, terminating at the top with a ball, and brought into connection with the inside coating by means of a light flexible wire or chain, attached to its lower end. Its capacity is approximately expressed by the formula

FIG. 388.

$$C = \frac{kS}{4\pi d}$$

where k is the dielectric constant for the glass, d its thickness, and S the area of either coating (supposed equal). If the jar be placed on an insulating stand and its inner coating be connected with a source of electricity, it will take only a comparatively small charge. If, however, its outer coating be grounded, the charge within may be increased many fold.

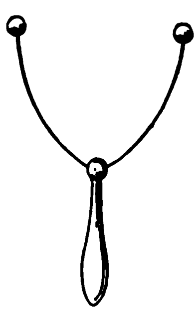


FIG. 389.

To discharge the jar safely an appliance called a discharger, Fig. 389, should be used. It consists of a pair of metal rods jointed together at one end, their other ends terminating in small balls. The jointed end is attached to an insulating handle. To discharge a jar the discharger is opened up and one ball placed in contact with the outside coat, while the other is brought up till it touches the ball of the jar. Just before touching, a spark will pass, indicating that the jar is discharged.

611. Residual Charges. Dielectric Strain. It frequently happens that after the first main discharge of a condenser one or two more progressively smaller residual ones may be obtained after brief intervals. This is a very important phenomenon and one that is very difficult to explain without taking account of the state of the dielectric. According to Faraday's view the dielectric between the plates of a charged condenser (and, in fact, a dielectric in any electrical field) is in a state of strain. Now we have already found (§ 179) that many substances show "elastic after-effects" when released from a state of strain; and it is a remarkable fact that such substances (and only they), when used as dielectrics in a condenser, give rise to residual charges. This is strong evidence for the view that electrification is essentially a state of strain in the dielectric or in the ether that permeates all matter.

The importance of the dielectric may also be shown in a striking manner by means of a "dissectable" Leyden jar, Fig. 390.

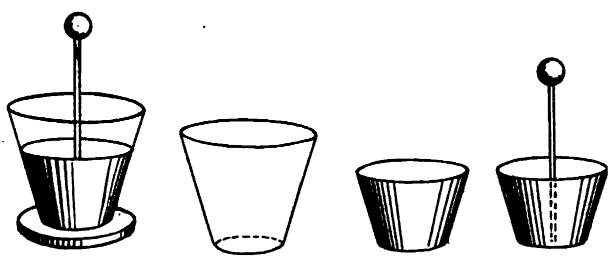


FIG. 390.

A represents the jar assembled and resting on an insulating pad. Suppose it to be charged, the outer coating being grounded during the process. The inner and outer coatings consist of rigid thin metal cups, fitting but not fastened to the larger glass tumbler which serves as the dielectric. If now the inner coat *D* be lifted out by non-conducting tongs, it is found on testing to be quite uncharged. Next *B* may be lifted free from the outer coating *C*, when the latter will be found also to be uncharged. If *C* and *D* even be touched together no spark passes. If now *C* be placed on the insulating stand and *B* be set in it and *D* within *B*, the whole being restored as at first, the complete jar will be found charged

as before it was taken apart, for on using the discharger a heavy spark may pass between the knobs. Thus what we speak of as charges on the plates is essentially a condition of the dielectric. Similar evidence is afforded by some experiments of Kerr. It is shown in the discussion of polarized light (§ 542) that when an isotropic body is placed between two crossed Nicol prisms, no light passes through the second prism. If, however, the body be put into a state of strain by any means whatever, it becomes temporarily double refracting, and the light incident on the analyzing Nicol is partly transmitted. Kerr tried the effect of a strong electrical field on carbon bisulphide through which a beam of polarized light was passing, the light being quenched by the analyzing Nicol before the electrical field was present. The presence of the field was always accompanied by a transmission of a part of the light through the second prism, the intensity being dependent on the strength of the field.

612. Dielectric Constants. In a number of our formulas we have already introduced a value k, which we have called the *dielectric constant* of the medium, and we are in a position now to explain a little more definitely how it is measured. Cavendish discovered, about 1775, that the inductive effect was different for different media and therefore we should expect the capacity of a condenser, to depend on the particular dielectric separating its plates (§ 591). The term *dielectric constant* was suggested by Maxwell and may be defined as *the ratio of the capacity of any condenser using a given substance as a dielectric to the capacity of a similar condenser having air for the dielectric.* This was called by Faraday the *specific inductive capacity* of the dielectric. Cavendish and Faraday measured experimentally the constants of various media. Some of the values obtained by these and other observers are given in the following table. As the values depend, in the case of solids at least, on the physical condition of the samples, as well as on the duration of the charge employed in the test, those in the table can be only approximate.

The measurements are made by using the various substances as dielectrics in condensers otherwise similar and comparing the resulting capacities. Many other methods have also been used.

Substance	*k*
Mica	8.0
Glass	6.0
Shellac	3.5
Sulphur	3.0
Petroleum	3.0
Benzine	2.5
Ebonite	2.5
Turpentine	2.0
Vaseline	2.0

Measuring Instruments. Electrical Machines.

613. Electrometers. These instruments, as the name indicates, are for measuring certain electrical quantities, notably potential differences. Some such instruments give a measure of the potential differences of two bodies in terms of other easily measured quantities, such as force, distance, and so on. These are called *absolute* electrometers. Other electrometers merely enable us to compare the potential differences of two bodies with that of two other bodies (such as the poles of a standard cell, § 665). The most valuable instruments of this sort were invented by Lord Kelvin and may be illustrated by describing two types, the *at-*

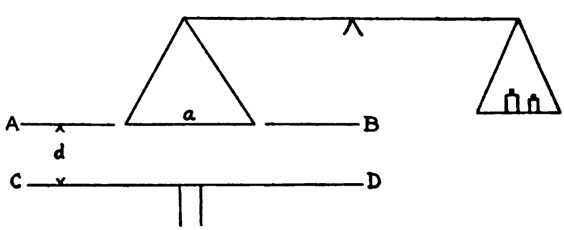


FIG. 391.

tracted disc absolute electrometer and the *quadrant electrometer,* which serves for comparing potential differences.

614. The Absolute Electrometer. Fig. 391 will make clear the general principle of the attracted disc instrument. *AB* and *CD* are two round conducting plates separated by a variable distance, *d,* which can be accurately measured. A central circular part, *a,* of *AB* is cut out, and while electrically connected with the rest of *AB* is movable up and down in the opening, being sus-

pended from one end of a balance beam. By means of sights adapted for the purpose, a can be adjusted accurately to the plane of AB when necessary. Practically AB and CD act as the two plates of an air condenser, the field being very uniform for the central part a. The ring part of the upper plate is retained only as a guard ring to insure the uniformity of the electric field between a and the corresponding part of CD. When no charge is present, a is exactly balanced in the plane of AB by the empty pan at the other end of the beam. When the plates are charged, weights are put in the pan just sufficient to balance the force F of attraction of a by CD. In order to use the instrument for measuring the potential difference between two points, AB (with a) is connected with one of the points and CD with the other, so that one of the plates will be charged positively and the other negatively. a then will be pulled down and the weights necessary to keep a in place will measure this attraction, F. From this the potential difference can be calculated.

The formula is derived as follows: The capacity of the parallel plates at a distance apart of d centimeters in air is

$$C=\frac{a}{4\pi d}, \quad (\S\ 608)$$

The energy required to charge the condenser is

$$E=\tfrac{1}{2}V^2C, \ (\S\ 596), \ = V^2a/8\pi d$$

in which V denotes the potential difference between the plates. The energy of the charged condenser is also equal to the work that the force F (§ 609) would do in drawing the plates together until they come into contact and their energy is reduced to zero. Hence $E=Fd$. Equating these values of E and solving for V, we get

$$V=\sqrt{\frac{8\pi d^2F}{a}}$$

in which d, F, and a are measured experimentally in C.G.S. units.

615. The Quadrant Electrometer. This instrument is shown in Fig. 392. The essential features are a circular thin metal box mounted horizontally on insulating pillars with a central vertical hole through it. The box is also divided into quadrants by mutually rectangular radial cuts. Suspended within the box by a very fine wire is a light flat conducting disc of the shape shown in

dotted line, Fig. 393. This "needle," as it is called, is kept charged to a high constant potential. The diagonally opposite pairs of quadrants are connected together and carefully insulated connections are brought to the outside of the case as shown. The indications of the instrument are noted by reflecting light from a fixed scale by means of a small mirror attached to the stem of the needle just above the case. When the pairs of opposite quadrants are at the same potential, the needle is so adjusted as to be placed symmetrically with reference to the quadrants, as shown in Fig. 393.

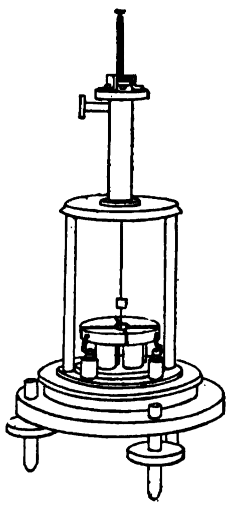


FIG. 392.

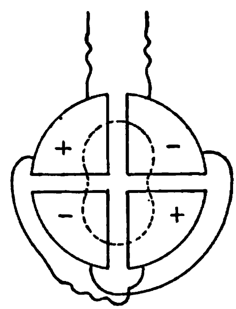


FIG. 393.

If, however, the diagonal pairs are at different potentials, the charged needle will rotate, owing to the attraction of one pair of quadrants and the repulsion of the other pair. If the quadrants be charged as indicated in the figure, the needle being positive, the latter will rotate clockwise till the couple due to the quadrants is just balanced by the torsion developed in the suspension. It can be shown mathematically that if the needle be at a very high potential compared with that of the quadrants, the deflection of the needle is directly proportional to the potential difference between the pairs of quadrants. Hence to calibrate the instrument the deflection produced by any known potential difference is noted and

from this the potential producing any observed deflection may be calculated.

616. Electrical Machines. The modern machines for producing electrification or electric charges continuously, and in far larger quantities than can be produced by simply rubbing two dissimilar bodies together and separating them, are the result of the evolution and of the aggregation of devices by many experimenters extending for nearly 150 years. The earlier ones were called *frictional machines,* because it was believed that the friction produced by the rubbing was an essential feature of the process. This, however, is not the case and the most efficient machines of the present day are active entirely by induction.*

617. The Electrophorus. The simplest induction machine is the *electrophorus* of Volta (1775). It consists of a metal pan, *a,* filled with a dielectric solid such as hard rubber or sealing wax, *b,* Fig. 394. A non-hygroscopic substance is preferable. Upon this is placed a smaller metal disc, *c,* furnished with an insulating handle *d.* To use the device, *c* is removed and *b* is electrified by stroking it vigorously with fur or flannel. If now *c* be replaced, the charge on *b* will act inductively on *c* attracting an opposite charge to the lower surface and repelling an equal similar charge to the upper surface. If now the upper surface be "grounded" for an instant, by touching with the finger for example, the repelled charge will escape and the disc *c* on removal will be found to have a free charge on it opposite in kind to that on *b*. *c* may now be discharged and replaced on *b* when the process may be repeated. Since *b* is a non-conductor and the contact between it and *c* is never good, its charge will remain practically undiminished and an indefinite number of charges may be drawn from *c*.

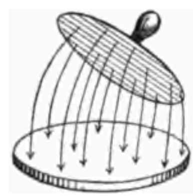


FIG. 394.

618. The Wimshurst Machine. By applying the electrophorus principle to a rotating plate, Holtz in 1864 constructed a generator which would produce a continuous electric discharge as long as the rotation was kept up. The inductive charging and

* For a description of the so-called frictional electric machines now obsolete in practice the student is referred to earlier books on the subject.

conductive discharging were brought about automatically. The early Holtz machine was improved by Toepler, Voss, Wimshurst and others.

The construction of the Wimshurst machine is quite simple and consists of two circular glass plates, each with a number of radial conducting sectors on the outside, and capable of rotating in opposite directions. The sectors act both as carriers and as inductors. Each plate has a diagonal conductor tipped at each end with a brush for making contact with the sectors as they pass. These conductors are at right angles with each other and inclined at about 45° to the vertical. There is also a collecting circuit with combs facing the plates, the capacity of which is usually increased by Leyden jars.

A very ingenious diagram suggested by S. P. Thompson, Fig. 395, will make the action of the machine clear. The sectors on the back plate are shown in diagram by the outer row, while the inner row represents those on the front plate. The two diagonal conductors are shown at *ab* and *cd*. Those carriers from which arrows extend are positively charged, while those towards which arrows point are negatively charged, the number of arrows indicating the amount of charge. As the front sector *a* passes into the position shown, it is influenced by the positive sector opposite to it on the back plate, but at the same time it is touched by the brush and a positive charge flows from *a* to *b* through the conductor, the sector, therefore, passing along with a negative charge. At the same time and for the same reason the sector at *b* is positively charged. It is thus clear that the front sectors advance from the brushes, the upper ones negatively and the lower ones positively charged. The diagonal conductor of the back plate intermediately produces the same result as can be followed

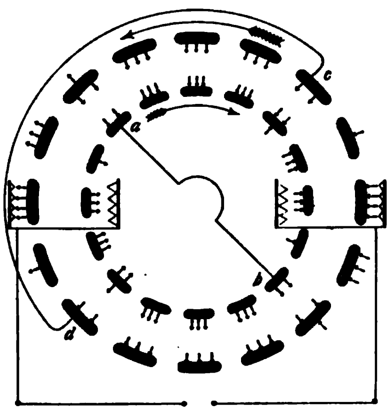


FIG. 395.

out from what has been said. The sectors of both plates, therefore, pass the combs of the horizontal circuit with like charges, positive on the left-hand side and negative on the right, thus charging them inductively by causing discharges from their points.

Prior to 1896, when the Röntgen rays were discovered, little use was made of these machines except for experimental purposes and no great power was sought for in their construction. Most of the larger machines consisted of only two plates about two feet in diameter capable of giving a nearly continuous threadlike discharge eight to ten inches long. Soon after Röntgen's discovery, however, it became the desire of all experimenters to develop as powerful rays as possible and these induction machines began to be constructed of large and numerous plates for exciting the "tubes" used in developing the rays. Machines consisting of many plates increased the quantity of the discharge per unit of time but not the difference of potential.

Atmospheric Electricity.

619. Identity of Lightning and the Spark. A lightning flash, like the spark from a condenser, is simply the disruptive discharge between opposed surfaces highly electrified. In each case the spark passes, when the accumulation of electricity is sufficient to cause the difference of potential between the surfaces to be too great for the dielectric strength of the intervening medium. When this occurs the dielectric breaks down and the discharge results.

FIG. 396.

FIG. 397.

Franklin in 1749 studied the subject and enumerated many points of similarity between the spark discharge and lightning. In the fall of 1752 he performed the now famous kite experiment, by means of which he was able during a thunder storm to draw charges from the clouds, and to perform the same experiments with them as with the sparks from his machines. In recent years the two phenomena have been compared photographically and the close similarity in the photographs of a lightning flash, Fig. 396, and of a spark from an induction machine, Fig. 397, furnishes

strong evidence that the two are identical. (The lightning photograph was taken by W. N. Jennings, of Philadelphia, in 1892; that of the induction spark by the author.) No attempt is made in the present work at an exhaustive account of the phenomena of atmospheric electricity or their causes. For such the student is referred to books on meteorology.

Just how the air becomes charged is yet somewhat of an open question though a number of explanations have been offered. The function of the clouds, however, seems clear. By their very act of forming they serve to concentrate into special regions the diffused electricity of the air and thereby gradually to cause a great potential difference between the clouds and the earth, or perhaps between two clouds themselves. Cloud formation, perhaps, consists in the gradual coalescing into larger globules of the water vapor at first more or less uniformly diffused in the air. Assuming each particle of the water vapor to have an initial small charge, it is easy to see how the combination of a large number of them into one drop must result in raising its potential to a very high value. If we consider 1,000 such particles, each with a charge which may be denoted by unity, to combine to form a single drop, the diameter of this drop will be only 10 times that of one of its particles, while its charge will be 1,000 times as great. Now the capacity of a sphere is proportional to its radius. Hence the potential will be raised 100 times by the combination, since $V = Q/C$. It is clear, then, since the initial charged particles are exceedingly small as compared even with one of the globules which go to form a cloud, that the high potential needed to account for a lightning flash may be explained in this way. The Aurora (Northern light) is due to electric discharges in the upper air but the full explanation is uncertain. The potential of the air is different from that of the earth. In fine weather it is positive and increases at a rate of 50 or 60 volts per foot of ascent. In stormy weather it is usually negative. Points on the surface of the earth may differ temporarily in potential and this difference gives rise to troublesome currents in telegraph wires called *earth currents.*

620. Protection From Lightning. Franklin in 1749 first suggested pointed iron rods attached to the higher parts of buildings and connected to the ground as a protection from damage by lightning. While public sentiment both for and against the use of *lightning rods* has at different times been strong, there seems to be no doubt that to use improperly constructed or ill-adjusted rods is far more dangerous to property than to do without them altogether. Lodge called attention a few years ago to the fact that it is not enough for a conductor to project above the building and terminate in a sharp point, that even more was essential than that it should be a good conductor of large cross-section and be well grounded. All these are necessary but not sufficient conditions. As a spark discharge is oscilla-

tory (§ 713), so is a lightning flash and such discharges have been shown to confine themselves to a rather thin layer of the conductor,—thinner in the case of iron than in that of copper. A large outside surface, then, is preferable to area of cross-section. In other words, a tube is just as good as a solid rod of the same diameter, and iron is better than copper. Furthermore, sharp bends and angles should be avoided and the conductor should get to ground by the most direct course. The conductor should be connected with a large plate or grid buried deep in moist earth. The water and gas pipes should be conductively joined together, but not to the lightning rod. It is well to ground *separately* the metallic trimmings of the building such as the roof, the gutters and the rain-spouts.

ELECTROKINETICS.

Continuous Conduction of Charges.

621. The Electric Current. Thus far we have considered under the head of Electrostatics only the phenomena of electric charges at rest, any motion being very temporary and consisting usually of almost instantaneous separation or union of two equal and opposite charges. Redistribution of charges on conductors due to induction has also been studied in the same connection. The phenomena of electricity in motion may be explained by supposing that whenever a conductor joins two points at different potentials a transfer of electricity called an *electric current* always takes place. a positive charge moving from the point at higher potential to that at lower, thereby determining the *direction* of the *current,* and a negative charge moving from a place of low to one of high potential. The current continues till both points and, in fact, all points on the conductor are at the same potential. It will therefore be only brief unless the two points are in some way *kept* at different potentials. In this case the current is continuous and if the potential difference be maintained constant, the *current strength* will be constant also, if the conductor is in no way changed. *Electrokinetics* therefore treats of electric charges in motion, or briefly of *electric currents.* The term current was adopted when it was supposed that electricity was some sort of a fluid and actually flowed through a wire just as a liquid flows through a pipe from a place of higher to one of lower level. More modern views on the subject agree, however, that the medium surrounding the conductor plays a very important part in all cases of electric flow. In fact

the surrounding dielectric is the medium by which *energy* is transferred from one place to another by so-called electric currents. This will be clearer when the magnetic fields accompanying currents have been demonstrated and discussed.

The magnitude of the current is the quantity of electricity which passes a given section of a conductor per second. If this quantity per second is constant, the conductor is said to be traversed by a current of constant strength. By quantity of electricity we here mean the sum of the positive transfer in one direction and the negative in the other direction. Now a flow of negative in any direction is equivalent in its effects to an equal flow of positive in the opposite direction. Hence for convenience, until we come to currents through fluids, we may now drop all reference to the negative stream and suppose it replaced by an equal positive stream in the opposite direction.

622. Voltaic Generators of Electric Currents. While it is possible to set up and maintain currents of electricity by means of an electrostatic machine, this is not the usual way. A much more convenient method was discovered by Volta. When two different metals are placed in dilute sulphuric acid, they are maintained at different potentials, so that, if they are joined externally by a wire, a current of electricity will flow along the wire. Such a current generator is called a voltaic cell. When the metals are zinc and copper the current through the wire is from the copper to the zinc, the former being at the higher potential.

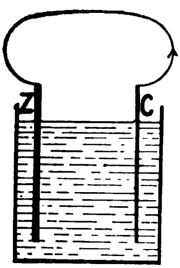


FIG. 398.

The electricity produced by this means differs in no essential respect from that produced by an electrostatic machine. For example, if one coating of a Leyden jar (or other condenser) be joined to one of the plates of the voltaic cell and the other coating to the other plate, the condenser will be charged; but the potential difference, as tested by an electrometer, will be found to be very small. It can, however, be greatly increased by the use of a large number of cells connected in a manner that will be described later.

In the account of the phenomena of electric currents that follows the use of the voltaic cell as generator will be assumed; but

a more complete account of the action of the cell can be given more conveniently at a later stage.

623. Magnetic Field due to a Current. The first to discover any relation between a magnet and an electric current was Oersted (in 1820). In that year he discovered that in general if a current be sent through any conductor near a suspended magnetic needle, the magnet is deflected from its natural position, that the amount of deflection depends on the relative positions and proximity of the two, and that the direction of the deflection depends on the direction of the current. Suppose the current in *AB*, Fig. 399, to be

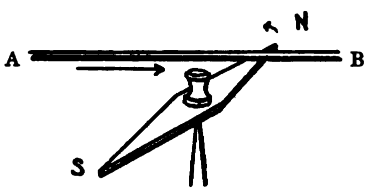


FIG. 399.

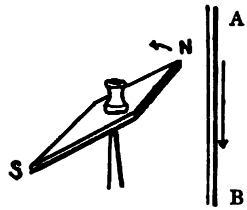


FIG. 400.

sent from south to north in a straight wire placed vertically above the magnetic needle, *NS*, which is initially in the magnetic meridian. There will be a deflection of the north pole towards the west. If the current be reversed so will the deflection, the *N* end moving now to the east. If the wire be placed below the needle instead of above it, the same change in the direction of deflection occurs as by reversing the current. Again if the relative positions of current and magnet be as shown in Fig. 400, *AB* being vertical, the *N* pole will turn west, being deflected east if *either* the current is reversed in direction in its present position *or* if *AB* be moved to the other end of the magnet. From these experiments, it is clear that if the needle be surrounded by a vertical loop or coil of wire in its own plane as in Fig. 401, the effect of the current in each part of the coil will be to produce a deflection of the needle in the same direction. We shall return to this point later (§ 637). The above experiments furnish strong evidence that there is a magnetic field associated with a current-bearing conductor. This is demonstrated graphically by leading a vertical wire which carries a current through a small hole in a horizontal glass plate, upon

which has been sifted iron fillings. If the plate be tapped gently several times, the filings will assume curved lines just as in similar experiments with magnets (§ 559). Fig. 402 shows the lines of such a field. In the present case the curves will be seen to be concentric circles, the wire passing through the center of the system. The direction of the lines may be determined by testing with a small suspended or poised magnetic needle. It will be found that, if the current be flowing in a direction from below the paper

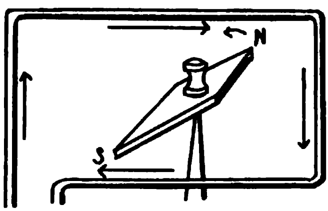


Fig. 401.

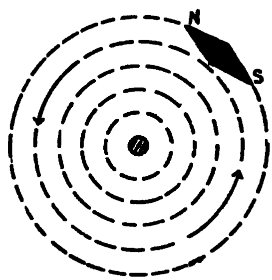


Fig. 402.

upward, the test magnet will rotate and place its axis nearly tangential to the circle passing through its center, with its *N* pole pointing in such a direction as to indicate that the lines have an anti-clockwise direction about the current. As an assistance in remembering the direction of deflection of a magnet when in the field of a current a number of rules have been devised. The simplest and most instructive of these is, perhaps, the rule given in § 71. To determine the direction of the lines of force around a current, recall the picture of an advancing right-handed screw; the current flowing in the direction the screw advances, the circular lines will be in the direction in which the screw rotates. This rule will be found very useful.

While in the above we have referred solely to ordinary or so-called *conduction currents,* it must be understood that, as discovered by Rowland, a charge of electricity on a moving body is equivalent to unit length of current *ev,* where *e* is the amount of the charge and *v* the speed of the carrier. Such a current is called a *convection current.* A third form of current, called a *displacement current* (Maxwell), is produced in a dielectric when the

latter is in a varying electric field. This form of current plays a very important part in the transmission of electric waves (§ —).

624. The C. G. S. Electromagnetic Unit of Current. We have already defined (§ 583) the C. G. S. electrostatic unit of electricity, and a C. G. S. electrostatic unit of current is flowing when the C. G. S. unit of quantity passes a given cross-section of the conductor per second. From these has been built up a system of electrical units known as the *electrostatic system*. It has been found, however, desirable in practice to make use of the magnetic effects just described as basis for a system of electrical units known as the C. G. S. *electromagnetic system*. We shall establish this system upon the definition of the unit magnetic pole, by defining first the electromagnetic unit of current strength. Suppose a wire carrying a current to be bent into the arc of a circle of one centimeter in radius, a unit positive magnetic pole being placed at the center. If the current be of such a strength that each centimeter length of it exerts a force of one dyne on the pole, the current is *one C. G. S. electromagnetic unit of current*. Thus we see that the static system of electrical units is based upon the definition of a unit charge, while the magnetic system is based upon the definition of a unit magnetic pole. As will appear later (§ 659) the unit of current used in practice is one-tenth of the C. G. S. unit and is called the *ampere*. If a unit current flow in a conductor, a unit quantity of electricity will flow across each section of the conductor per second. If the current be one ampere the quantity will be one *coulomb*, which is therefore the practical unit of electrical quantity. It is one-tenth of the C. G. S. electromagnetic unit of quantity.

625. Electromotive-force. Newton defined force as anything that tends to move matter. By analogy then that which moves or tends to move electricity may be called *electromotive-force*. But it is not force in the mechanical sense, since what it moves is not matter. The analogy is, however, a useful one and for clearness we shall pursue it a little farther.

A difference of level will cause a flow or current of water. Similarly, a difference of potential causes a current of electricity. A pump may be used to raise water to a higher level, so that it will descend in the form of a current. To this corresponds a

voltaic cell, which raises electricity from a lower potential at the zinc plate to a higher potential at the copper plate, and so causes a current when the plates are connected by a wire. But a difference of level is not necessary for a flow of water; a rope drawn through a horizontal pipe containing water will produce a current. Similarly, we shall see later that a current of electricity can be produced (by induction, § 683) in a circuit in which there are no differences of potential.

These analogies, while suggestive, must not be pressed too far. They will, however, serve to illustrate the statement that there are various forms of E. M. F. The simplest is a difference of potential, $V_1 - V_2$, at the two ends of a wire of the same temperature throughout. In this case we may speak of the potential difference as the E. M. F. acting on the wire. If, however, the wire be divided and a voltaic cell be inserted, the total electromotive force, E, between the ends of the wire is the algebraic sum of the potential difference, $V_1 - V_2$, of its ends and the electromotive force, e, of the cell, the latter being *the potential difference of the plates of the cell when they are not joined by an external conductor,* that is, when the current is not completed. For the present (until we take up Thermoelectricity) we shall use E in the sense here defined.

The C. G. S. electromagnetic unit of potential difference or of electromotive-force exists between two points of a conductor, if one erg of work must be done on an electromagnetic unit quantity of electricity, in moving it from the point at the lower potential to that at the higher. This is far too *small* for practical purposes and therefore a practical unit called the *volt,* 10^8 times as large, has been adopted.

626. Ohm's Law. The strength of a current through a given conductor depends on the magnitude of the E. M. F. which produces the current. The relation which the magnitude of the current bears to that of the E. M. F. is extremely simple and was first discovered by Ohm. *The magnitude of the current in a conductor is proportional to the magnitude of the E. M. F. or potential difference* (if this be the only form of E. M. F. present), or, stated as a formula,

$$I = \frac{E}{R}$$

where R is a constant that is always the same for a given conductor (in a constant physical state) but different for different conductors. The value of R for any conductor is called the ***resistance*** of the conductor. The resistance of a conductor is, therefore, defined as ***the constant ratio of the E. M. F. applied to a conductor to the current which it produces.***

For many purposes it is more convenient to write Ohm's Law in the form

$$I = C \cdot E,$$

where C is a constant for the conductor that carries the current. Since this constant C is proportional to the current, I, which the body conducts, it is appropriately called the ***conductance*** of the body, and it may be defined as the ratio of the current through a conductor to the E. M. F. From the above equation it follows that $C = 1/R$.

Ohm's Law is of fundamental importance and has been tested with great care. Chrystal found that the resistance of a wire for infinitely weak currents did not differ from its resistance for very strong ones by 10^{-10} per cent.

If a part of a circuit contains a voltaic cell of electromotive force e, and if the whole resistance of that part of the circuit be R, we may, as in § 625, substitute $V_1 - V_2 + e$ for E and write Ohm's Law in the form

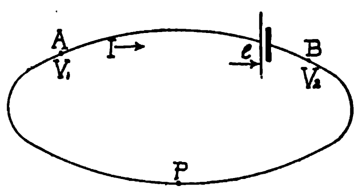


FIG. 403.

$$I = \frac{V_1 - V_2 + e}{R}.$$

Since this formula applies to any part of a circuit, it must apply to the whole circuit, reckoned from some starting-point around the circuit to the same point. In this case $V_1 = V_2$ and

$$I = \frac{e}{R}$$

where e stands for the E. M. F. of the cell (or cells) contained in the circuit and R for the total resistance of the circuit.

627. Definition of the Ohm. In accordance with the resolutions

of the International Electrical Congress, Chicago, 1893, and adopted by act of Congress in 1894 the legal definition of the ohm is that unit of resistance which "shall be what is known as the international ohm, which is ***substantially equal*** to 10^9 units of resistance of the centimetergram-second system of electromagnetic units, and is represented by the resistance offered to an unvarying electric current by a column of mercury at the temperature of melting ice, 14.4521 grams in mass, of a constant cross-sectional area, and of the length of 106.3 centimeters."

If I is measured in amperes and E in volts, then R is expressed in the corresponding practical units of resistance. The practical unit of resistance is called an ***ohm***. Since an ampere is 10^{-1} absolute electromagnetic unit and the volt is 10^8 such units, it follows that the ohm must be equivalent to 10^9 C. G. S. electromagnetic units of resistance.

628. Laplace's Law. A very convenient expression for the magnetic intensity produced by any small part of a current was suggested by Laplace, and was verified by Ampère by numerous ingenious experiments, which are described in more complete treatises. Laplace's law can be most easily expressed by a formula. Let dx be a short part of a circuit carrying a current I. If a line of length r, drawn from a point P to the middle of dx, makes an angle a with dx, the magnetic intensity, dH, at P due to the current in dx is

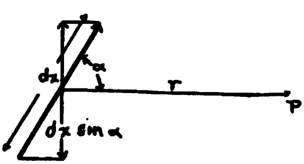


FIG. 404.

$$dH = \frac{I dx \sin a}{r^2}$$

Since $dx \sin a$ is the projection of dx on a plane perpendicular to r, it is evident that the intensity, dH at P due to dx depends only on this projection. This, taken along with the right-handed-screw rule, also gives us the direction of the force at P due to the current in dx.

629. Strength of Magnetic Field Due to a Straight Current. Consider a part AB of a straight conductor, Fig. 405, carrying a current I and a point P so located that PA is perpendicular to AB. Then the intensity

at P at right angles to PA due to the current in AB will evidently be

$$H=\int_0^l \frac{Ir\cdot dx}{(r^2+x^2)^{\frac{3}{2}}}=\frac{I}{r}\left[\frac{x}{(r^2+x^2)^{\frac{1}{2}}}\right]_0^l=\frac{I}{r}\sin APB$$

If now AB be extended indefinitely, sin APB will approach unity as a limit. Moreover, if the lower half of the conductor be also added, then the force at right angles to PA will be doubled anl in general the strength of the field at P will be

$$H=\frac{2I}{r}$$

Fig. 406 represents a cross-section of a current and its field in which A is the conductor, the direction of the current being from below upwards. Then the field will have an anti-clockwise direction as indicated by the

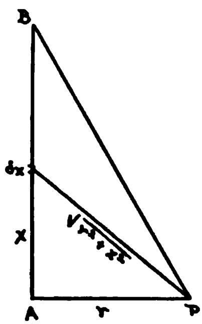


FIG. 405.

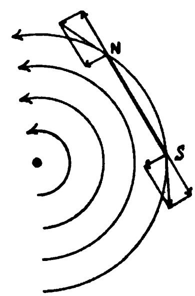


FIG. 406.

arrows. Suppose a small magnetic needle NS brought into the field. It will place itself as shown (§ 623). The same line of force will pass through both N and S, and the force acting on each may be resolved into two components, Fig. 406, one parallel to the magnet's axis and the other perpendicular to it. The first pair of components will be equal and *opposite,* but the other pair, though equal, will be in the *same* direction and will tend to move the magnet as a whole nearer to the wire. This fact will be noticed when mapping the field with iron filings (§ 623); as the surface is tapped the circles will gradually contract in size, the filings all moving toward the current.

Work Done in Taking a Pole of Unit Strength Around a Current. Since the force at P, Fig. 404 is $2I/r$, to move a unit pole in the circumference of radius r completely around the current will require work equal to $2I/r \times 2\pi r$ ergs, or $4\pi I$ which is *independent of the distance.* Hence the work of conveying a unit pole completely around a current by *any path* and back to starting point is constant and equal to $4\pi I$.

631. Strength of Field due to Circular Current. Let us consider first the case in which P is at the center of the circle. Then all the elements of the current being at a distance r from P,

$$H=\frac{lI}{r^2}=\frac{2\pi rI}{r^2}=\frac{2\pi I}{r}$$

the direction being that of a perpendicular to the plane of the circle and through P. If there are n turns instead of one,

$$H=\frac{2\pi nI}{r}$$

Let P be located, not in the plane of the coil, but on the axis through its center perpendicular to that plane (Fig. 407). Consider the effect of a small element dl of the current at A. The force at P due to this is

$$dH=\frac{Idl}{d^2}$$

and its direction is perpendicular to d and is represented by Pb. This force Pb may be replaced by two mutually perpendicular components, Pa and Pc, Pa being perpendicular to the axis OP and the other along this line. If the same reasoning be applied to the element of current at B, diametrically opposite from A, its influence at P may be represented by Pa' and Pc', Pa' being equal and opposite to Pa, but Pc' being equal and in the same direction as Pc. As the whole circuit AB can be divided up into pairs of diametrically opposite elements to which the same reasoning can be applied, it follows that the Pa components all sum up to zero, while the Pc components are all additive in the direction OP. Since $Pc=Pb\sin\alpha$, the total effect due to the whole circuit is

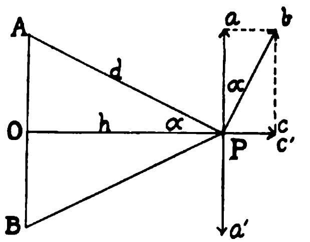


FIG. 407.

$$H_p=\Sigma Pc=\int dH\cdot\sin\alpha=\int\frac{I\sin\alpha}{d^2}dl=\int\frac{Ir}{d^3}dl=\frac{2\pi r^2 nI}{(r^2+h^2)^{\frac{3}{2}}}$$

As a special case this reduces to $2\pi nI/r$ when $h=0$, *i. e.*, when P moves up to O, a result obtained above.

Electromagnetic Effects.

632. Action Between a Magnet Field and a Current Field. We have just seen (§ 623) that if a magnet be brought near a con-

ductor bearing a current, it is acted on by a force causing it to place itself in a particular way. Now, since action and reaction are equal and opposite, if the magnet be not allowed to adjust itself into a position of equilibrium with reference to the current field, there must be a force tending to move the conductor so that by *its* motion the relative positions of the magnet and conductor may be properly adjusted.

Suppose the magnetic field to be rather strong and uniform (Fig. 408 *A*), and let the conductor extend perpendicular to it, the current flowing from above downward (*B*). The current lines are circles and have a clockwise direction. Now suppose the two fields to be superposed (*C*), evidently the field will be strengthened above the wire and weakened below it. The direction of the resultant field will be changed at all points except at those in a meridian plane *AB* containing the wire. At each point, as at

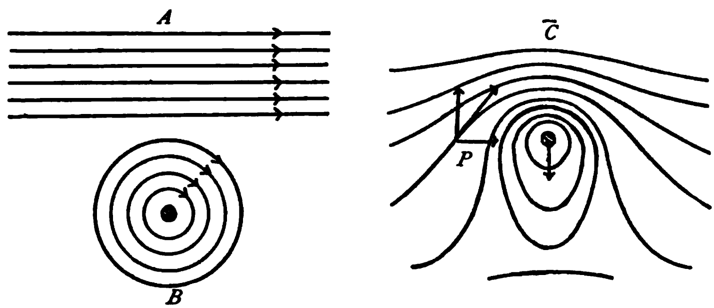


Fig. 408.

P, its direction and magnitude will depend on the two components, that of the magnet field and that of the current field, which together determine the resultant field at the point. It follows from this, and from the tendency of lines of force to contract, that there will be a resultant force tending to move the wire in the direction *AB*. A rule suggested by Fleming enables one to remember the direction of this force. If the forefinger of the left hand be held extended in the direction of the magnetic field, and the middle finger point in the direction of the current, the direction of the outstretched thumb shows the direction of the resultant force. This is evidently in the direction perpendicular both to the current and to the field. It was shown with great care by Am-

père that the force developed on a conductor of length l carrying a current of strength I perpendicular to a field H, is lIH. If the conductor be not perpendicular to the field, the force will be proportional to the cosine of its inclination to this direction.

From this it is clear that the current strength in C. G. S. units is numerically equal to the force in dynes with which a unit magnetic field tends to move one centimeter length of the conductor carrying the current.

633. The Solenoid. In § 631 we discussed the strength of the magnetic field at various points on the axis of a circular current. Suppose now we have a number of such circles all on the same axis and arranged so that a current flows similarly through all of them. The helix, Fig. 409, will represent closely such an arrange-

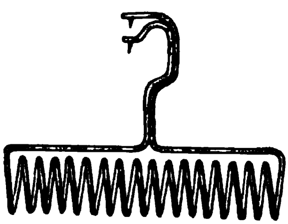


FIG. 409.

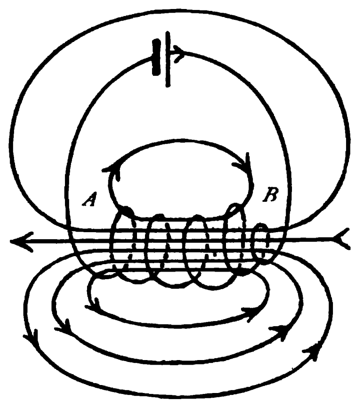


FIG. 410.

ment, which is called a *solenoid*. The points at the top enable one to suspend it from two mercury cups to which current connections may be made in such a way as to give the whole free movement about a vertical axis. The resultant lines of the field within the solenoid are parallel to the axis. Fig. 410 represents such a field and its similarity to that of a permanent magnet is at once apparent; in fact a solenoid has all the properties of a magnet. The face A will be its north pole, and B its south, as is readily seen by applying the right-handed screw rule to find the direction of the magnetic force. Two such solenoids will react on each other exactly like two permanent magnets, and any suspended

solenoid carrying a current will set its axis parallel to the magnetic field in which it is placed.

The similarity of action of a solenoid and a magnet strongly suggests that magnetism may be due to solenoidal currents, and this was first proposed by Ampère. This theory, known as Ampère's theory of magnetism, supposes that the molecules of a magnetic substance have closed electric currents circulating about them without resistance and therefore without diminishing. To magnetize such a body the molecules must be rotated by a magnetic field (§ 634) so that their axes are in the same direction, the strength of the resulting magnet depending on how perfectly this is brought about.

634. Electromagnetism. If a paramagnetic substance be placed within a solenoid carrying a large current, it will become magnetized by induction to an extent depending on its permeability (§ 565). For example, if a rod of soft iron a little longer than the helix be used, a powerful magnet will result, but only while the current is passing through the helix. If the current be interrupted the magnetism will almost entirely disappear. This combination is called an *electromagnet* and has a very wide application in both industrial and experimental apparatus. In order to utilize both poles the soft iron core is often bent in the form of a U and each leg is surrounded by a separate coil of silk-covered wire wound, for convenience, on a spool, the winding being such that if the core were straightened out the two coils would form a continuous one. Fig. 411 shows the usual type, the soft iron piece across the poles being called the *armature;* this serves to complete the magnetic circuit through the iron. Other things being equal, the induction per unit cross-section is proportional to the product of the current strength, I, and the number of turns of wire per unit length, n, and in C. G. S. units (§ 624) is equal to $4\pi nI$ if there be no cord in the solenoid, or $4\pi\mu nI$ if the core have a permeability μ. As μ may have a value of 2000 or more, and I may be made as great as we please, it is clear how superior an electromagnet may be to a permanent steel magnet. If I is in amperes $B = 4\pi\mu nI/10$.

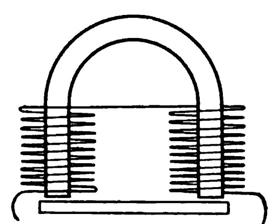


FIG. 411.

A proof of the above may be briefly stated as follows: Let Fig. 412 represent a longitudinal section of a coil having n turns of wire per unit length, through which flows a current I C. G. S. units. Let $abcd$ be a small rectangle with sides parallel and perpendicular respectively to the axes of the solenoid, ab and cd being of length a, so that the number of coils that pass through $abcd$ is na. The work done in taking a unit-pole around the path $abcda$ is therefore, $4\pi naI$ (§ 629). If the intensity along ab is H the work done in this part of the path is Ha. The force along bc and that along da are zero (if there were any such forces they would, in any case, be equal and opposite). Hence the work along bc and da is zero. If the helix be sufficiently long so that its ends are far enough away the force along cd may be neglected compared with that along ab (this is also indicated by the absence of lines of force near cd, compare Fig. 410). Hence the total work is Ha which must therefore equal $4\pi naI$. Therefore $H = 4\pi nI$.

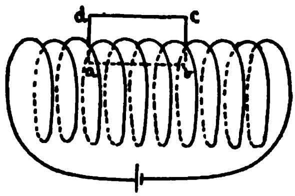


FIG. 412.

635. Energy Relations. Attention is again directed to the surrounding medium as the seat of a part of the energy expended in an electromagnetic circuit. Since an effect at a distance cannot result instantaneously from any cause, we must assume that the field around a wire opens out gradually though with great velocity. Now a magnetic field contains energy since it can do work in moving a magnetic pole. Hence for a brief interval after closing a circuit a part of the energy supplied to the circuit by the generator is expended on the field, and the current does not instantly rise to its full value even in a straight wire. Again when the current is broken this field closes in again on the conductor and yields up its energy, thus maintaining the current for a brief time.

636. The Telegraph Sounder. One of the simplest and commonest applications of the electromagnet is its use in ordinary telegraphy. The *sounder*, Fig. 413, consists of an electromagnet with an armature pivoted so that it may swing to and fro in front of the poles with a short amplitude between stops. The receding motion is produced by an adjustable spring. A device called a key for closing and opening the circuit, which contains also a source of current, is placed in series. The con-

ducting wires may be as long as is consistent with the E. M. F. at one's disposal. When the key is closed the armature is drawn down with a click; when it is opened the spring reacts and another click follows. The signals which indicate the letters and numbers consist of various combinations of short and long intervals between the down and up clicks, a short interval of closing being

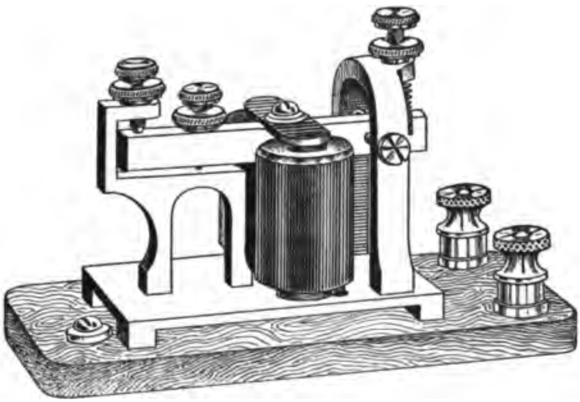


FIG. 413.

called a *dot* and a longer one a *dash*. A short open circuit interval is called a *space*. Operators become very expert at reading these signals by ear. This system was devised in 1836 by Morse and his code with little change is still in general use.

In practice various *relay* devices are introduced by means of which the distance between the first sending and the final receiving stations may be as great as desired. The feeble current received from a distant station operates an electro-magnet, which closes a fresh circuit with a relay supply of generators. This station thus at once becomes a new sending station, the process being automatic.

637. The Tangent Galvanometer. The application of the principles of §§ 623, 631 enables us to construct one type of a class of instruments known as galvanometers which are used for measuring current strengths. The simplest form of galvanometer consists of one or more circular turns of insulated copper wire of radius a decimeter or more, the plane of which is placed in the

magnetic meridian, Fig. 414. The insulation is effected by covering the wire with one or more layers of silk. At the center of the coil a light magnetic needle a couple of centimeters or less long is poised on a sharp point or suspended by silk or quartz fiber. When no current is passing, the needle, being in the magnetic meridian, is in the plane of the coil. If now a current passes through the coil, a field is developed perpendicular to its plane and the needle is deflected till the moment of the deflecting couple is just balanced by the moment of the earth's field. It will be noticed that, as the needle turns, the arm

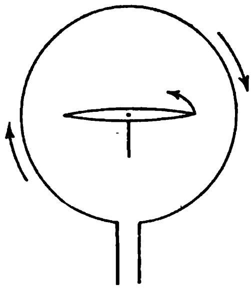


FIG. 414.

of the deflecting couple is decreasing while that of the restoring couple is increasing; at some angle of deflection then, ϕ, these two couples must balance and the needle will come to rest. From § 630 the field at the center of the coil is $2\pi nI/r$, and $2\pi nI/r \cdot m$, then, is the force which acts on each pole of the needle of magnetic strength m. The moment of the deflecting couple is $2\pi nIm/r \cdot l \cos \phi$ if ϕ is the angle of deflection and l the length of the needle, Fig. 415. If now H is the strength of the earth's field, the restoring or balancing couple due to it is equal to $mH \cdot l \sin \phi$. Equating these and cancelling common terms, we get

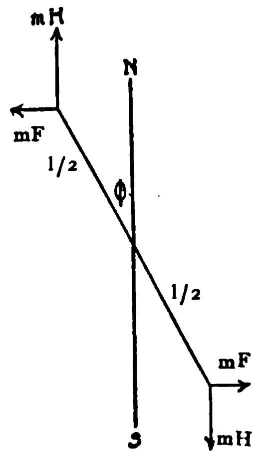


FIG. 415.

$$I = \frac{rH}{2\pi n} \tan \phi$$

If we substitute G for $2\pi n/r$, we get

$$I = \frac{H}{G} \tan \phi = K \tan \phi$$

from which it follows that the strength of the current is propor-

tional to the tangent of the angle of deflection produced by it. For this reason the instrument is called a *tangent galvanometer.* Since G is a function of the dimensions only of the particular instrument used, it is known as the *galvanometer constant,* and $H/G = K$ is called the *reduction factor* of the instrument. The current is expressed in the above formula in C. G. S. electromagnet units. In amperes

$$I = \frac{10H}{G} \tan \phi$$

In order that the magnetic field at and near the center of the coil shall be as near uniform as possible, so that the needle as it rotates shall have its poles always acted on by the same force, Helmholtz suggested an instrument composed of two parallel vertical coils of the same radius in series or parallel circuit, placed at a distance apart equal to the radius of each. The needle should be located half way between the coils on their common axis. In this form the needle need not be so short compared with the radius of the coil.

As the value of H varies from place to place as well as from time to time at the same place, it follows that, for the measurement of quantities involving H, its value should be determined at the time and place that the observations are made. In the comparison of currents, however, at any one time, the relation of the tangents of the deflection angles only need be considered.

638. The Thomson Galvanometer. The instrument just described is designed to measure currents of moderate strength which are met with in laboratory practice, say from 0.1 of an ampere to 10 amperes. Far weaker currents, however, frequently have to be studied and measured, and for this purpose a different type of instrument is designed. Of the very sensitive instruments, one developed many years ago by Lord Kelvin (then Sir William Thomson) and known as the Thomson galvanometer is the most important. One of its early uses was as a receiving instrument for detecting the weak currents used in the Atlantic cable when laid in 1865. No other instrument known at the time could have served the purpose. It employs two devices not yet described, viz., the *astatic system* of magnets and a *mirror* and *scale* for reading the deflections. The former increases greatly the sensitiveness of

the instrument and the latter enables the observer to read the deflections much more accurately than by other means. It is clear that, by using in a galvanometer two needles of equal magnetic moments but exactly reversed in direction, a system may be obtained upon which the resultant action of the earth's field will be zero. Since such a system has no directive tendency, it is called *astatic* and will remain at rest in any direction. Fig. 416 represents the arrangement of such a system in a galvanometer. The two needles *ns, s'n'* are rigidly connected and relatively reversed. A coil surrounds the upper one and sometimes also the lower, and, if the action of the current on each be traced, it will be found that the effect is additive and that both needles conspire to rotate the system in the same direction. For this reason alone the sensitiveness is great. In order still more to increase the sensitiveness the needle system is suspended by a very fine filament of silk without torsion. If, under these circumstances, the system were exactly astatic the least current would develop field enough to rotate it nearly 90°, the angle depending on the slight torsion of the fiber, and no measurement would be possible though the instrument would be an infinitely sensitive galvanoscope. To obviate this, the earth's field, the effect of which is eliminated, is replaced by an artificial field produced by a weak bar magnet, *NS*, Fig. 416, placed above the instrument and adjustable as to height and angle. Since *NS* is nearer to *ns* than to *s'n'* its directive influence on the combination is not balanced and *ns* controls the system. Clearly by such means the sensitiveness can be changed at will. Practically the astatic system is never exactly balanced and so the earth's field has always some little directive influence. By the governing magnet *NS*, however, in combination with the earth's field the resulting field may be adjusted to any desired strength. Another function of the control magnet is to enable the instrument to be used in any position.

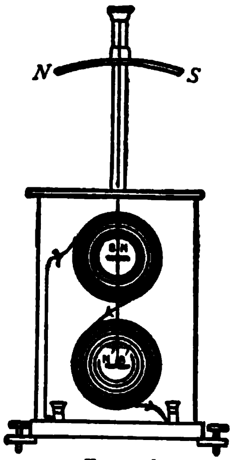


FIG. 416.

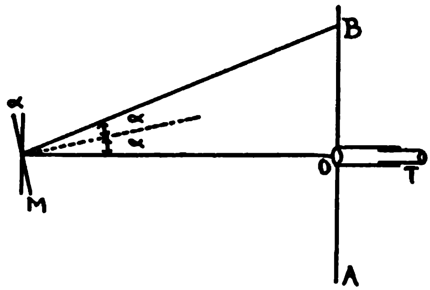


FIG. 417.

The mirror M referred to is very small and light, sometimes only a few millimeters in diameter. It is usually attached to the system by cementing to its back the upper magnet, composed of a bit of magnetized watch spring. By observing the deflection of a reflected beam of light, directed properly upon the surface of the mirror, the angle turned through by the system can be calculated, the latter being half as great as the deflection of the beam of light (§ 390). Fig. 417 will make the method clear. AB is a graduated scale well lighted. M is the mirror, its equilibrium position being parallel to the scale. T is a telescope with eyepiece cross-hairs, the axis of which is perpendicular to the mirror. Usually the zero (o) of the scale is just under the telescope, so that before M rotates one looking into the telescope may see the zero line in the middle of the field. If now M rotate a few degrees, other divisions of the scale will come into the field of the telescope, and when M comes to rest the scale reading is noted. Suppose it is b cm. (BO in the figure), then if MO is c cm. (often about one meter) we may write $b/c = \tan 2\alpha$, in which α is the angle turned through by the mirror M; therefore

$$\alpha = \tfrac{1}{2} \tan^{-1} \frac{b}{c}$$

Since α is usually a very small angle, it is often accurate enough to assume that any two currents are proportional to the corresponding values of b, *i. e.*, to the deflection readings.

639. The D'Arsonval Galvanometer. Another method of construction is to reverse things and make the magnet fixed and large, while the coil is movable and small. The arrangement is shown in Fig. 418 and is known as the D'Arsonval type of galvanometer. N and S are the poles of a strong permanent magnet and between these, with its plane parallel to the field, the coil is suspended by means of a delicate torsion conductor through which the current is led to the coil. To the other and lower end of the coil is attached a spiral spring, which runs

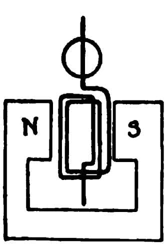


FIG. 418.

vertically downward to the bottom of the instrument, and which is drawn just tight enough to hold the coil in place and yet allow it to rotate freely about a vertical axis. Within the coil is usually placed an iron cylinder to increase the induction. A mirror is fixed to the suspension just above the coil and a telescope and scale is used. When a current is passed through the coil the straight vertical parts of the latter tend to move perpendicular to the lines of force of the magnetic field of the magnet (§ 632). This is opposed by the torsion of the suspension. For small angles the current is very approximately proportional to the deflection it produces. While the D'Arsonval type is not as sensitive as a Thomson instrument, it has the great advantage of not being appreciably affected by outside influences. Iron or steel in the vicinity has little or no effect on it, and it may be made practically dead-beat, *i. e.*, so that it comes to rest without vibrating. This is due to the reaction induction currents set up in the coil due to its motion in the magnetic field. The direction of all such currents is always so as to oppose the motion (see Lenz's Law, § 685).

640. The Ballistic Galvanometer. This is an instrument designed for measuring a quantity, Q, of electricity discharged from a condenser or other charged body. It may be made of either the Thomson or the D'Arsonval type and the latter is now the more common, it being so constructed for this purpose that its "damping" is small and its period long. The current maintained is extremely brief—in fact has usually ceased to flow before the needle or coil has moved appreciably, provided the period be sufficiently long. The relation between the quantity Q and the "throw" of the needle, as measured by scale divisions seen in the telescope, is approximately expressed by the simple equation $Q = kn$ in which k is a constant easily found by observing n, the scale reading, for a known charge Q and substituting these values in the equation.

This principle is often used in comparing the capacities of condensers. If two or more condensers be charged successively to the same potential and then discharged through the same ballistic galvanometer, the quantities of electricity Q in each case will be proportional to the corresponding scale readings. But we have learned that $C = Q/V$ (§ 591); therefore since V the potential is constant, the capacities are also proportional to the scale readings.

641. Voltmeters. Any galvanometer that has a coil of very high resistance takes very little current, especially if it be joined between two points of a circuit otherwise connected through a very much smaller resistance, *i. e.*, if it be used in shunt circuit (§ 648). In this case, however, the little current it does take will be proportional to the potential difference between the points to which it is joined, and with a properly calibrated scale it is made to read volts and is then called a *voltmeter*.

642. Direct Reading Instruments. For practical purposes in industrial use a number of remarkably accurate empirical instruments have been devised for measuring both current and potential difference. They are called respectively *ampere-meters,* or *ammeters,* and *voltmeters.* The quantity to be measured is indicatd at once by the position of a pointer over a scale graduated to read directly either amperes or volts. The Weston meters are standard instruments of this class. The principle is that of the D'Arsonval galvanometer, the movable coil turning on hard steel pivots on agate bearings. There is a fixed cylinder of soft iron within the coil to intensify the magnetic field. The motions are remarkably deadbeat. The zero position of the coil is governed by non-magnetic spiral springs.

RESISTANCE.

643. The **resistance** of a conductor has been defined as the ratio of the E. M. F. applied at its ends to the current strength resulting. Resistance depends on the kind of material, length, shape, etc., of the conductor. It depends also on the temperature, in some materials increasing and in some decreasing with a rise of the temperature.

The resistance of a wire for electricity is in certain ways similar to the resistance of a pipe for the flow of water in it. The conductance in both cases is proportional to the size, *i. e.*, to the cross-section, while the resistance in both cases is proportional to the length. This, while a useful analogy, must not be pushed too far for the resistance of a pipe depends on the strength of the current through it.

Electrical resistance may also be likened to friction between solids. Thus a copper or silver wire is a *smooth* conductor for electricity while an iron or a lead one is a conductor ten times as *rough* electrically. It has already been pointed out that no substance is a *perfect* conductor and none a perfect non-conductor or insulator.

644. Resistivity. It is found that the resistance of a uniform wire is proportional to its length and is inversely proportional to its cross-section, *i. e.*

$$R = \frac{\rho l}{a}$$

in which ρ is a constant for any one substance. Evidently $\rho = R$ when both l and a are unity, or ρ is numerically equal to the resistance of a rod 1 cm. long and 1 sq. cm. sectional area and it is called the *resistivity* or *specific resistance* of the substance.

RESISTIVITY, ρ_0, OF PURE METALS (SOFT AND ANNEALED) AT 0°, AND MEAN TEMPERATURE COEFFICIENTS BETWEEN 0° AND 100°.*

Metal.	ρ_0 (*C. G. S. units*).	a_{0-100}
Platinum	10,917	0.00367
Gold	2,197	0.00377
Palladium	10,219	0.00354
Silver	1,468	0.00400
Copper	1,561	0.00428
Aluminum	2,665	0.00435
Iron	9,065	0.00625
Nickel	12,323	0.00622
Tin	13,048	0.00440
Magnesium	4,355	0.00381
Zinc	5,751	0.00406
Cadmium	10,023	0.00419
Lead	20,380	0.00411
Carbon (Incandescent Lamp)	4×10^6	

645. Effect of Temperature on Resistance. The resistance of a conductor changes with a change of temperature. The resistance of pure metals and of most alloys is increased by a rise of temperature.

The following curves, Fig.* 419, show in the case of several metals the variation of resistance with temperature from —200° or lower to +200°. Resistances are expressed in C. G. S. units of the electromagnetic system. It will be noticed that the general trend of convergence is toward a point of zero resistance at a

* Dewar and Fleming, *Phil. Mag.*, Vol. 36, p. 299.

* From Nichols and Franklin's *Physics*, representing the results of Dewar and Fleming.

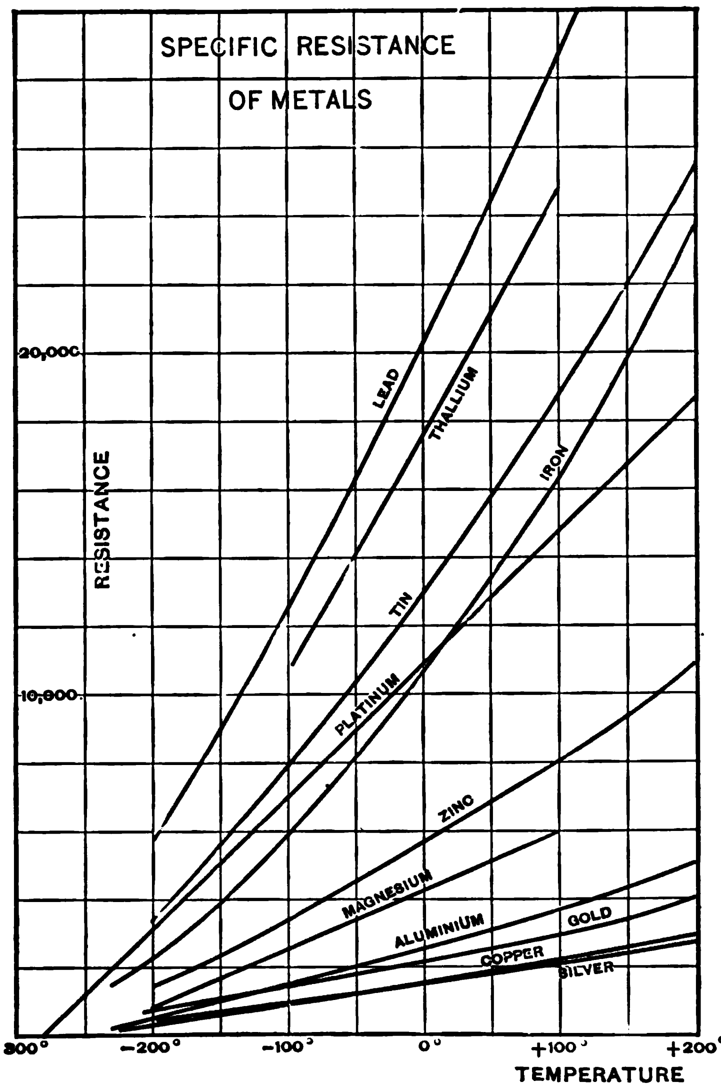


FIG. 419. (Nichols and Franklin, after Dewar and Fleming.)

temperature between —250° and —300°. As —273° is the most probable position on the centigrade scale of the absolute zero, it follows that at or near the absolute zero of temperature the resistance of all pure metals would be zero. As the curves shown have a very long radius of curvature, it follows that over comparatively

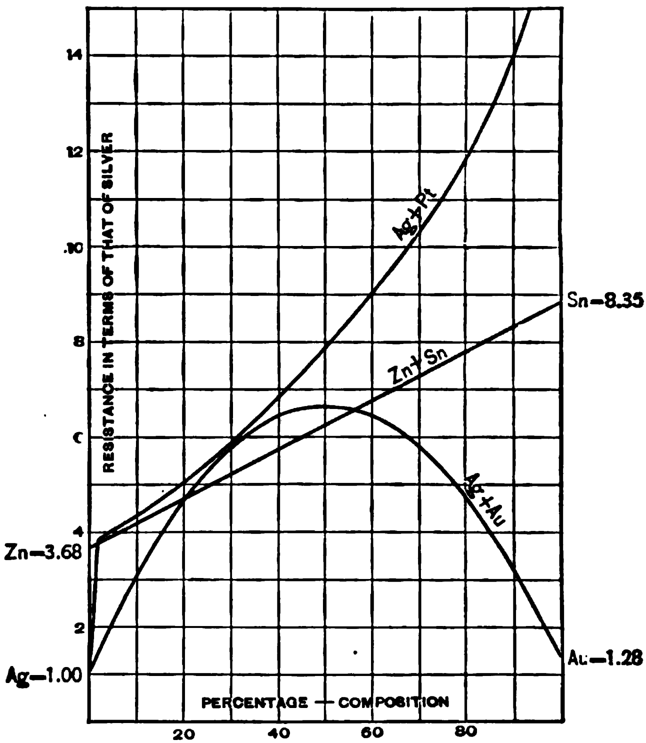


FIG. 420.

small ranges of temperature the increase in resistance is nearly proportional to the increase in temperature, or

$$R_t = R_0(1 + at)$$

in which R_t is the resistance at any temperature $t°$, R_0 that at some standard temperature as 0°, and a is the *temperature coefficient* of the particular metal. For all pure metals the temperature

coefficient is found by experiment to be approximately .00366, or equal to the coefficient of expansion of a perfect gas. From the diagram it seems that the temperature coefficient of platinum is most nearly constant over a wide range of temperature. As seen by the diagrams and as has been found by Dewar the curves rise somewhat as absolute zero is approached.

646. Resistance of Alloys. As regards their electrical resistance alloys or mixtures of pure metals are divided into two very distinct classes. Alloys containing lead, tin, zinc, or cadmium have resistivities which can be calculated from those of the pure metals and from the composition. Alloys of most other metals have resistivities which are much greater than would be anticipated, and the temperature coefficient is less than in the case of either constituent. This latter fact is of great value in the construction of *standards* and of resistance coils for laboratory or commercial use. For example, if 12 parts of nickel, 84 parts of copper and 4 parts of manganese be mixed together, an alloy called *manganin* is formed having a temperature coefficient very nearly zero. Fig.* 420, constructed from the results of Matthiesen, shows by three examples the anomalous behavior of alloys.

647. Standards of Resistance. Since the comparison of resistances can be made much more accurately than that of any other electrical quantities, it is very desirable that standards should be available which do not change with time or with change of temperature. Such have been made possible by the use of insulated manganin wire, doubled and wound on a spool, the two ends of the wire thus appearing at the same end of the spool. These are then soldered to stout terminals of copper rod, bent in such a way as conveniently to be inserted into mercury cups for connection.

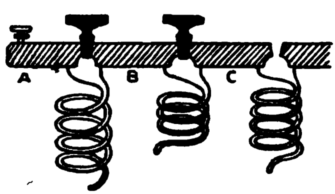


FIG. 421.

For laboratory use sets of coils wound as above described are mounted in boxes, with their terminals so connected on the outside of the cover that by removing a metal plug from between two blocks to which the terminals of a coil are joined, that coil

* From Nichols and Franklin's *Physics*.

is brought into the circuit. Combinations of this sort are known as *resistance fluxes*. Fig. 421 shows the manner of winding these coils and joining them to the blocks. Boxes may be bought with coils varying in resistance from 0.01 ohm to 100,000 ohms.

648. Resistance of Divided Circuits. There are two distinct ways of joining conductors together. They may be joined end to end, so that the current must go through them one after another, the strength of the current being the same in each one. Conductors so arranged are said to be *in series*, and the total resistance of the circuit is equal to the sum of resistances of all the parts. If, however, a number of conductors as *a, b, c,* Fig. 422, are joined with their coresponding ends together, as at

FIG. 422.

A and *B*, they are said to be joined *in parallel* or *in multiple*, and a current passing from *A* to *B* will divide itself between the three conductors, the one offering the least resistance taking the largest part of the current. If the potential difference between *A* and *B* is *E*, then the drop of potential from *A* to *B* will be the same along each of the three paths, *a*, *b*, and *c*. If the resistances of the branches be r_1, r_2 and r_3 respectively, and the currents in them i_1, i_2 and i_3, then by Ohm's law

$$i_1 = E/r_1, \; i_2 = E/r_2, \text{ etc.}$$

Since after the current has become constant there is nowhere an accumulation of electricity the total current flowing from *P* to *Q* must be the sum of the partial currents, or

$$I = E(1/r_1 + 1/r_2 + \text{etc}) = E/R$$

Hence the branch conductors between *A* and *B* behave as though they were replaced by a wire having a resistance *R* such that

$$\frac{1}{R} = \frac{1}{r_1} + \frac{1}{r_2} + \text{etc.}$$

Hence when conductors are arranged in parallel the conductance

of the system is equal to the sum of the conductances of the several parts. Since $E = RI$ it follows that

$$i_1 = \frac{R}{r_1} I, \quad i_2 = \frac{R}{r_2} I, \text{ etc.}$$

In such divided circuits as occur between A and B, any one of the paths a, b or c is called a *shunt* to the others. Shunts are frequently used between the terminals of a galvanometer or other electrical device, for the purpose of diverting a part of the current which otherwise would all pass through the instrument. Thus by knowing the resistance of the instrument and of the shunt separately, we can at once calculate what fraction of the total current passes through the galvanometer and is measured. Sensitive galvanometers are usually furnished with appropriate shunt coils, so adjusted that say only 0.1 or 0.01 or even 0.001 of the total current is actually passing through the instrument. The instrument is thus protected from frequent danger of injury. In this case the whole current, I, is expressed in terms of the measured current, i, and the resistances of the galvanometer, g, and of the shunt, S, by the formula

$$I = \frac{S+g}{S} i \quad \text{(see above)}$$

so if $S = \frac{1}{99}g$, for example, then $I = 100i$.

649. Kirchhoff's Laws. The calculation of the distribution of currents in complex branched circuits is greatly facilitated by two laws derived by Kirchhoff. In the combination circuit AB, Fig. 422, we saw that the current I was equal to the sum of the currents in the branches, or

$$i_1 + i_2 + i_3 - I = O$$

from which follows at once the first law: *The algebraic sum of all the currents that flow towards a point in any system of conductors is zero.*

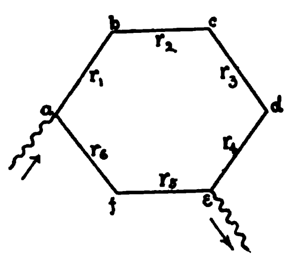


FIG. 423.

Again, suppose we have any *closed* circuit of conductors which does not include an E. M. F., as *abcdefa*, Fig. 423. The current in ab may be expressed by the equations,

$$i_1 = \frac{V_a - V_b}{r_1} \quad \text{or} \quad i_1 r_1 = V_a - V_b$$

That in *bc* also may be written

$$i_2 = \frac{V_b - V_c}{r_2} \quad \text{or} \quad i_2 r_2 = V_b - V_c$$

and similarly for the other parts of the closed circuit taken in order. If now we add these equations, we get

$$i_1 r_1 + i_2 r_2 + i_3 r_3 + i_4 r_4 + i_5 r_5 + i_6 r_6 = 0, \text{ or } \Sigma ir = 0$$

If such a closed circuit does contain an E. M. F., then the potential differences will not add up to zero, but to Σe, where each e must be given a positive sign if it is in the direction taken as positive for the current, and the equation becomes

$$\Sigma ir = \Sigma e$$

which may be considered an extension of Ohm's law. From this result Kirchhoff's second law is stated as follows: *In any closed circuit, no matter how complex, the sum of the products obtained by multiplying the resistance of each part by the strength of the current flowing through it is equal to the sum of the E. M. F.'s in the circuit.*

650. Fall of Potential Along a Current Bearing Conductor. If the two ends of a uniform homogeneous conductor be maintained at a constant difference of potential, a uniform current will

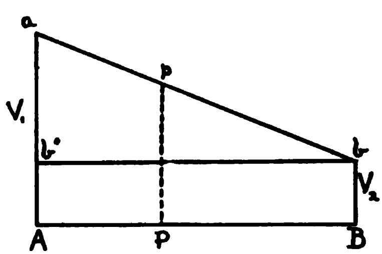


FIG. 424.

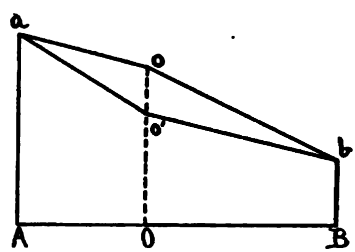


FIG. 425.

flow from one end to the other and the fall of potential along the conductor will be uniform per unit of length. This follows at once from Ohm's law, $E = IR$, and from the fact stated that R is the same per unit length throughout the wire, I having the same

value at all points. We may represent this by a diagram, Fig. 424, in which AB represents the length of the uniform conductor. From A erect an ordinate, Aa, representing the potential V_1 at A, and at B, Bb, representing the potential V_2 at B. Now since the drop is uniform, the potential locus for all points between A and B will be a straight line from a to b, and the potential at any point P will be represented by Pp. The whole fall from A to B is $V_1 - V_2$ and is represented by ab'. Suppose now the conductor between two points be composed of two uniform sections, AO and OB of different lengths and of different resistances per unit length, Fig. 425. Then ao represents the gradient of potential in the part AO if this part has a less resistance per unit length than OB, while ao' will represent it if AO has a greater resistance per unit length than OB. Similarly ob or $o'b$ shows the gradient in OB. As before and in general the fall of potential between any two points is proportional to the resistance between those points.

The fall of potential along a resistance may be used to com-

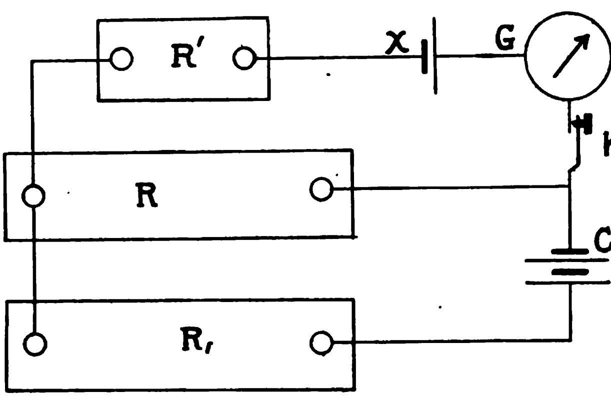


FIG. 426.

pare small electromotive forces by the following method due to Rayleigh. The plan of the apparatus is as shown in Fig. 426. R and R_1 represent two adjustable resistances of 10,000 ohms each as a maximum. These are joined in series with a cell C having a somewhat higher E. M. F. than those to be compared. During

the adjustment of the resistances R and R_1 a total of 10,000 ohms must always be kept in circuit. The cells x and y to be compared are joined one at a time in a shunt circuit around R, including a higher resistance R', a sensitive galvanometer G, and a key K. The cell x should be so joined as to oppose the potential difference between the terminals of R. By transferring resistance between R and R_1 a balance may be secured so that on closing K no deflection results in G. Then the E. M. F. of x is equal to the potential fall through R. Again, on replacing x with another cell y, a similar balance can be secured. Then if R_x and R_y are the corresponding resistances in R in the two cases, $E_x : E_y = R_x : R_y$.

651. Wheatstone's Bridge. This consists of an arrangement of conductors the purpose of which is to measure resistance. In the diagram, Fig. 427, P and Q are maintained at a fixed difference of potential, that at P, for instance, being the higher. As the fall of potential from P to Q is the same whether one considers the path PaQ or the path PbQ, it is clear that for any point a on PaQ there can be found some point b on PbQ, such that the fall between P and b shall bear the same ratio to the whole drop as that between P and a does. In other words, wherever a may be chosen, a point b can always be found at the same potential as a. This can easily be done in practice by joining a to a sensitive galvanometer and testing for a point b, so that no current is indicated by the instrument. When b is so located, then the ratio of the fall of potential in Pa to that in aQ is $r_1 : r_2$ and the ratio of the fall in Pb to that in bQ is $r_3 : r_4$. Hence it follows that $r_1/r_2 = r_3/r_4$. In practice r_1/r_2 is a fixed and known ratio. If now r_3 be an unknown resistance, the experiment consists in adjusting r_4 till a balance is obtained, that is, till no current flows through the galvanometer. When this adjustment is made the unknown resistance is at once calculated from the equation,

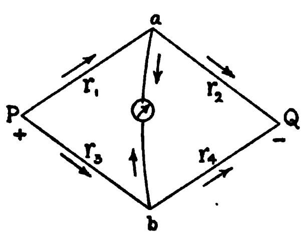


FIG. 427.

$$r_3 = r_4 \cdot \frac{r_1}{r_2}$$

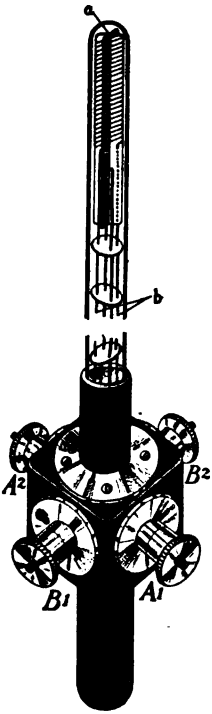


FIG. 428.

652. The Platinum Thermometer. The fact that the electrical resistance of a pure metal can be expressed as a simple function of its temperature is made use of in constructing a very accurate thermometer for temperatures both above and below the limits within which the mercury thermometer is practicable. For this purpose platinum is used, its curve being practically a straight line. The precautions to be taken and the best method of construction have been thoroughly investigated by Callendar and Griffiths and their arrangement is shown in Fig. 428. A wire of pure platinum is doubled and wound on a thin mica frame *a* and the whole inclosed in a hard annealed glass or porcelain tube. The ends of the wire at *b* are welded to two thick copper leads ending in binding screws A_1A_2 in the hard rubber cap which is fitted to the top of the tube. Since the resistance of the lead wires will also change with the temperature while it is only the resistance of the platinum coil that we wish to obtain, a compensating device is introduced consisting of a like pair of leads attached to the binding screws B_1B_2, and joined together at their lower ends. All four wires terminate at about the same place in the instrument and run close together so as always to be at the same temperature. When the resistance between A_1 and A_2 is measured, B_1B_2 is introduced into the circuit of the adjacent arm of the bridge (§ 651) so that any variation of resistance of the lead wires to A_1A_2 will be balanced by the same variation in the dummy leads B_1B_2. The platinum thermometer gives excellent results up to about 1000°.

ELECTROLYTIC CONDUCTION.

653. Eectrolysis. In the preceding pages have been described the phenomena of electric charges at rest and also in motion, the latter taking place in solid conductors. In the case of such conduction no final change has occurred either in the composition or physical state of the conductor. It is true that if a metallic wire be heated by the current to a temperature above its fusing point the metal will melt, but it will solidify again as soon as the *heat* is withdrawn. The electricity by itself causes no change in the conductor. Certain liquids, as for example mercury, behave ex-

actly in the same way, while some others, as petroleum and turpentine, act just like solid dielectrics; that is, they do not conduct at all. Still a third class of liquids, however, notably a very large number of the chemical compounds, either fused or in solution, behave in a manner quite different from anything that has been before considered. The transfer seems to be more like an *electrical convection,* and the substance is decomposed. For this reason the process is called *electrolytic conduction* or *electrolysis,* and the substance is called an *electrolyte.* If, for example, we dip into a solution of HCl, contained in the glass beaker *A,* Fig. 429, two platinum or carbon strips, p_1p_2, and include these in the electric circuit of an appropriate generator, *B,* a current will flow in the circuit showing that the liquid in some way has transferred the electricity. The terminal of the strip p_1 which is joined to the wire by which the current enters is called the ***anode,*** while the terminal by which the current leaves the *electrolytic cell* is called the ***kathode.*** The two together are called the ***electrodes.*** As has been mentioned the liquid is decomposed into two substances, either elements or compounds, one appearing at the anode plate and one at the kathode plate. That part of an electrolyte which appears at the anode is called the ***anion,*** while that which appears at the kathode is the ***kation.*** All these electrolytic terms were originally suggested by Faraday.

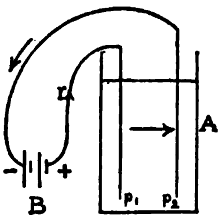


FIG. 429.

It is not always the case that the initial products of electrolysis are given up free at the electrodes, since frequently secondary reactions occur between them and the materials of the electrodes or perhaps the unchanged electrolyte. In the case of HCl, Fig. 429, the hydrogen will be given off in bubbles at the kathode plate, p_2, while the chlorine will appear at the anode, p_1. The process is as if the hydrogen ions traveled through the electrolyte towards p_2, carrying positive charges, while the chlorine ions moved towards p_1 with negative charges. These positive and negative charges are given up to their respective electrode plates. If a solution of H_2SO_4 be the electrolyte, hydrogen is again the kation, while the anion SO_4 combines with a molecule of H_2O, forming again H_2SO_4 and giving free oxygen at the anode. The ratio of

the volumes of the H and the O is 2 : 1, or the same as is necessary to form water. For this reason it was long supposed that water was the electrolyte, the acid being added simply to make it conducting. This, however, is not the case, for, as is explained above, the H_2SO_4 is the substance electrolyzed.

654. Faraday's Laws. The generalizations of electrolysis were stated in 1833 by Faraday. They are two:

I. *The mass of an electrolyte decomposed is proportional to the quantity of electricity which passes through it.*

II. *The mass of any substance liberated by a given quantity of electricity is proportional to the chemical equivalent of the substance.*

The number of hydrogen atoms which is necessary to form a stable molecule by combining with an atom or group of atoms is called the *valence* of the atom or group. In considering the molecules HCl, H_2O, H_2SO_4, it is clear that the valence of Cl is one, that of O is two, and that of the group SO_4 is two. The ratio of the atomic mass of an atom, or of the sum of the atomic masses of a group of atoms to the valence is called its *chemical equivalent*. Thus the chemical equivalent of O is 8. We may write as a combined expression,

$$m = \gamma Q = \gamma It$$

in which m is the mass of a given substance, say hydrogen, γ is a constant depending on the particular substance, and the other letters have their usual meaning. From the equation it appears that γ is equal to the mass of the substance set free by a unit quantity of electricity, or by unit current in one second; it is called the *electro-chemical equivalent,* and is proportional to the chemical combining mass of the substance. Since, then, if the value of γ be known in any one case, its values for other substances can be calculated from their chemical equivalents, it is desirable to know the values of γ for a few familiar substances. The measurement can be made with great accuracy in the cases of the solutions of certain metal salts, such as sulphate of copper or nitrate of silver. If, for example, in the electrolytic cell, Fig. 429, the platinum strips be replaced by strips of silver and the solution be one of silver nitrate, the silver from the solution will attach itself to p_2, and if this electrode be weighed before and after a known current I has passed for a time

t, the amount, *m*, of silver deposited can be determined, and γ calculated from the equation

$$\gamma = \frac{m}{It}$$

Experiments of this kind have shown the value of γ for silver to be 0.001118 grams per coulomb. We may with nearly equal accuracy find γ for copper and for various other metals.

655. The Charge on a Single Ion. Since the atomic mass of silver is 107.94 and its valence is 1, while the atomic mass of hydrogen is 1 and its valence is also 1, the electro-chemical equivalent of hydrogen is

$$\gamma = \frac{0.001118}{107.94} = 0.000010357 \text{ grams per coulomb}$$

Therefore, since $m = \gamma It$, in any particular case the mass m_s of any substance liberated will be given by the equation

$$m_s = 1.0357 \times 10^{-5} qIt$$

in which q is the chemical equivalent mass of the ion in question.

As one coulomb of electricity will separate 0.001118 gram of silver, it will require 107.94/0.001118 or 96,550 coulombs to separate the chemical equivalent in grams of silver or 1 *gram equivalent*. Now, by recalling Faraday's second law, it is clear that it will require the same number of coulombs, viz., 96,550, to liberate 1 gram equivalent of any substance. Since, passing through an electrolyte, 96,550 coulombs always liberate 1 gram equivalent of each ion, it follows from the convection idea, which assumes the positive charges to be carried by the kations in the direction of the current and the negative charges by the anions in the opposite direction, that the charge carried by the chemical equivalent of each substance must be the same. Since the electrochemical equivalents of univalent substances are proportional to their atomic or to their molecular masses, the charge carried by each individual substance must be of the same magnitude for all univalent substances. If (e) is the value of this charge in coulombs,

$$(e) = 96{,}550(m)$$

in which (m) is equal to the mass in grams of a single ion of

hydrogen. If the atom of hydrogen is the mass of its ion, then, since the best authority gives for its mass the value of about 10^{-25} gram, (e) is about 10^{-20} coulombs.

The charge carried by a divalent ion is $\pm 2(e)$, of a trivalent ion, $\pm 3(e)$, etc. If the charge (e) be measured in electromagnetic units instead of in coulombs, then

$$(e) = 9{,}655\,(m)$$

or the ratio of charge to mass in the case of a hydrogen ion in liquid electrolysis is approximately 10^4. This result has an important bearing on the subject of the discharge of electricity through gases (see Radioactivity).

656. Theory of Electrolysis. Some of the facts of electrolysis have been known for more than a century and attempts have been made from time to time to construct a theory which would explain the phenomena consistently. If, for example, copper sulphate be electrolyzed between copper plates, the solution around the kathode continually becomes weaker, while that around the anode becomes more concentrated. If platinum plates be used instead of the copper ones, the solution about both grows weaker but the kathode side much faster than the other. This fact was explained by Hittorf by the *migration of ions,* by assuming that different ions move through the liquid at different speeds, and Kohlrausch later proved this experimentally and showed further that each kind of ion has its own specific speed for a given gradient of potential in a given liquid, independent of the particular molecule or group of atoms from which it originated. Hydrogen has a higher speed than any other ion for the same rate of fall of potential through the solution. For this reason it follows that acids which are hydrogen compounds have a greater conductivity than their salts in solution. Since by careful investigation it has been shown that both Ohm's law (§ 625) and Joule's law (§ 667) hold in electrolytes, all the energy of the current when traversing the liquid must be used in heat as in a solid metal conductor and none of it can be used in doing chemical work in separating molecules into ions. From this it follows that the electrolyte must always contain at least some ions in a free state, a part of these charged with $+(e)$ and others with $-(e)$. It has already been noted that perfectly pure water is almost a non-conductor and it is also true that pure sulphuric acid is practically a non-conductor. A mixture, however, of these forming a solution is an excellent electrolyte. We describe this by saying that in the act of solution we have a process of *ionisation,* or the production of a condition by which a compound becomes a conductor, and we conclude that the mole-

cules of sulphuric acid and of water are in a different state when in solution than when separated; they are said to be *dissociated* by the act of solution. It has been shown that for very weak solutions most if not all the molecules are in a state of dissociation, while for stronger ones only a portion of them are in a state of free ions. It is probable, too, that separation and recombination is constantly going on, all the ions changing partners frequently. The undissociated molecules at any instant play no part in the transfer of the electric charges. The dissociating influence of water may be due to the high specific inductive capacity, k, of water, for it has been shown (§ 583) that the attraction between charges is inversely proportional to k; and if the natural molecule is composed of two parts equally and oppositely charged they will be more easily separated in a medium for which k is large. Other properties of solution also support the dissociation theory. In brief, then, electrolytic conduction consists in the convection of positive and negative charges by two sets of ions moving in opposite directions through the liquid, one set giving up its charges to the kathode plate, the other to the anode. The speed of the ions is proportional to the potential gradient in the liquid but for any given gradient each kind of ion has a specific characteristic speed independent of its source.

657. Polarization. The cause of the motion of the ions is the difference of potential between the electrodes, but, when they have arrived at the electrodes, the potential difference may not be great enough to cause the charge to leave the ion. If the charge does go to the electrode, it plays its part in neutralizing the charge there supplied by the generator, and the uncharged ion combines with others and the substance is in part recomposed. If the charge does not leave the ion because of insufficient potential difference, the ions collect on the oppositely charged electrode and set up there an opposing electromotive force which is called *polarization*. Further, if the generator be removed and the circuit again completed by joining the two wires, an inverse current, gradually becoming zero, will pass through the galvanometer. This inverse current is due to the polarization built up at the electrodes during the initial passage of the direct current, and the effect of electrical polarization is much like that of a reaction spring or cushion in mass dynamics.

658. The Voltameter or Coulometer. Faraday suggested the use of the electrolytic cell to measure current strengths. This assumes a knowledge of the electrochemical equivalent of the substance deposited from the solution. For ordinary currents

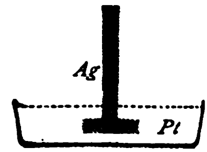


FIG. 430.

probably silver is the best material to use. The best kathode is a clean platinum dish, upon which the silver is deposited, Fig. 430, and in which is suspended a disc of pure silver supported by a silver rod without solder. The electrolyte is composed of a 15 per cent. solution of pure silver nitrate in distilled water and is placed within the dish. The current is calculated from the formula (§ 654),

$$I = \frac{m}{0.001118t}$$

in which I is measured in amperes, m in grams and t in seconds. It may here be noted that the legal definition of the ampere by It may here be noted that the legal definition of the ampere in accordance with the resolutions of the International Electrical Congress, Chicago, 1893, and adopted by act of Congress in 1894 is that "the unit of current shall be what is known as the international ampere, which is one-tenth of the unit of current of the centimeter-gram-second system of electromagnetic units, and *is the practical equivalent* of the unvarying current, which, when passed through a solution of nitrate of silver in water in accordance with standard specifications, deposits silver at the rate of 0.001118 of a gram per second."

659. Electroplating. There are many applications of electrolysis in the industrial arts, the most important of which perhaps is *electroplating*, or the process by which a thin layer of a metal is deposited from a solution over the surface of a body by means of the electric current. For *electrotyping* copper is used and is deposited on the surface to be copied. The copper layer is then removed and is filled at the back by pouring in melted type-metal. Such plates are far more durable than the original type.

660. Secondary Cells. It has already been pointed out (§ 656) that a polarized electrolytic cell is capable of maintaining an inverse current for an appreciable time, if the generator be removed and the wires be joined together. As a result of long investigation on this subject, started by Planté (1860) and continued by Brush, Faure and others, the commercial *secondary* or *storage cell* has been developed, the function of which is to absorb electrical

energy from a generator, the energy being rendered potential by electrolysis, and then to return that energy, or a large part of it, in electric current at any desired time or place. The electrolytic process is called *charging* the cell, and in subsequently using the cell it is said to be *discharged*.

In its most practical shape at the present time the storage cell consists of a series of perforated or grid-like plates of an alloy of lead with a small percentage of antimony. Into the perforations is pressed a paste composed of lead sulphate made by mixing litharge and red lead with sulphuric acid. These prepared plates are immersed in a strong solution (20 to 25 per cent.) of sulphuric acid. During the process of charging, the hydrogen ions pass to one of the electrode plates, react upon the paste of lead sulphate, depositing metallic lead in a spongy form and forming sulphuric acid which goes into the solution. At the other plate, the anode, appear the SO_4 ions which in their reaction upon the sulphate, form lead oxide, PbO_2 and sulphuric acid according to the following equation:

$$PbSO_4 + SO_4 + 2H_2O = PbO_2 + 2H_2SO_4$$

The sulphuric acid adds itself to the solution and the plate is oxidized. It is seen, then, that the action of the charging current has been to reduce the paste on the kathode plate to metallic lead, and to peroxidize it on the anode plate. The cell is fully charged when free hydrogen is being given off from the kathode, as this indicates that the active material has all been reduced. The cell may now be used as a generator, giving a current in the direction opposite to that of the charging current. Its E. M. F. is about 2.2 volts, which decreases as the discharge goes on. It is detrimental to the cell to discharge it so long that its E. M. F. becomes less than 1.8 volts. The efficiency is almost 80 per cent. The storage capacity of storage cells is usually rated in ampere-hours, one ampere-hour being 3,600 coulombs. The necessary great weight of the storage cell in its present form is a great disadvantage, interfering with its portability. Also the plates are apt to deteriorate owing to the expansion and contraction of the active material during charge and discharge, thus making it liable to fall out from the perforations and short circuit the cell. The current from such a source, however, is very constant and is exceedingly convenient for many purposes.

CHEMICAL GENERATORS.

661. General Methods. So far in the development of the general subject we have said very little about the practical means

of generating continuous electric currents of any appreciable magnitude. There are two general methods, the chemical and the dynamic. The former will be considered first. In treating of electrolysis we have seen that chemical energy may result from the expenditure of electrical energy, and the converse should be expected, namely that chemical energy may be changed into electrical energy.

662. The Volta Effect. About the year 1786 an Italian physiologist, Galvani, while working with some detached frog's legs near an electrical machine in operation, noticed that at the moment of a discharge the legs would give a slight jerk. Subsequently he observed that similar specimens would behave in the same way, when blown by the wind against an iron railing from which they were hung by a copper hook. The observation was the subject of a spirited discussion between Galvani and a relative of his, Volta by name, who could not accept Galvani's explanation that the effects were due to the physiological action between the tissues of the animal. Volta held that the effects were due to the electrification brought about by the contact of dissimilar substances, and he was thus led to the discovery that when a plate of zinc and one of copper are placed in dilute sulphuric acid and joined by a wire, a current of electricity traverses the wire. There must therefore be differences of potential at some or all of the contacts of zinc, acid and copper and the algebraic sum of these is the E. M. F. that causes the current.

Great difficulty has been found in determining the potential differences at these contacts. Volta believed that the contact between the two metals played the most important part, and, as a result of his experiments, he arranged metals in a series such that each was positive to the next. Lord Kelvin also adopted this view as a result of experiments in which a needle like that in the quadrant electrometer (§ 615) was suspended above two metals in contact. Others have found that the series is entirely different, if these experiments are performed with the metals in gases other than air. In fact, all such experiments merely measure the E. M. F. of a cell consisting of two metals and a gas. Surface polish, absorption of gases, and previous contact with other metals affect the apparent potential differences at contact of metals. As a result of the extensive labors of numerous workers, it is now generally accepted that the potential differ-

ences due simply to the contact of metals are at least extremely small. Of the potential difference at the contact of metal and liquid there is no doubt. It can be measured in various ways.

663. Voltaic Cells. In the storage cell described in § 659, the potential energy was derived by the expenditure of electrical work upon the cell. If, however, the substances be initially assembled in the construction, so that the initial effect is a chemical action resulting in an electric current, the cell is known as a primary or voltaic cell. If such a cell contain dilute sulphuric acid and a zinc anode plate, the H_2SO_4 will separate, the H_2 being replaced by an atom of zinc because in the combination of zinc with SO_4 more energy is set free than is required to separate H_2 from SO_4. In this change, therefore, there is an amount of energy developed which appears as that of an electric current. When the zinc anode and a copper plate for the kathode are immersed in the dilute acid, and connected by a copper wire, the first effect is a volta difference of potential between all the contacts, the algebraic sum of which taken around the circuit is 1.08 volts. This is the initial E. M. F. of the cell. But the current obtained from the cell cannot be calculated from this E. M. F. In fact the current is observed to die away very rapidly. Hence an opposing difference of potential must have developed somewhere. This is found to be at the junction between the copper and the hydrogen, to be in fact due to the polarization by hydrogen explained in § 657. Many uses have been found for cells which would deliver a fairly constant current for considerable time, and an early effort was made to correct the defects due to polarization. If possible the hydrogen is removed either mechanically or chemically. The latter method is the better and often consists in supplying sufficient oxygen for the hydrogen to combine with.

In the *Bichromate* cell this is done by adding potassium bichromate which furnishes the oxygen necessary to oxidize the hydrogen.

The *Bunsen* cell consists of a large zinc anode in H_2SO_4 solution, within which is a porous earthen cup containing a carbon rod immersed in strong nitric acid. As the hydrogen ions pass freely in through the walls of the cup the positive charges are given to the kathode while the hydrogen is quickly oxidized by the nitric acid. Such a cell has an E. M. F. of about 1.7 volts.

The *Daniell* cell is perhaps the most constant of any in general use and requires very little attention. Its kathode is a copper plate in a solution of $CuSO_4$, the zinc anode being immersed in dilute sulphuric acid which is contained in a porous cup standing in the $CuSO_4$ solution. The chemical reaction may be stated as follows, $Zn + ZnSO_4CuSO_4 + Cu = 2ZnSO_4 + 2Cu$. A form of this is the *gravity* cell (Fig. 431) which does away with the porous cup, the copper plates surrounded by the copper solution being in the bottom of the containing vessel. The zinc plate is hung near the top of the vessel, the solution of H_2SO_4 surrounding it. The difference in density between the two solutions helps to keep them apart. There is no tendency to mix diffusely while the cell is in action.

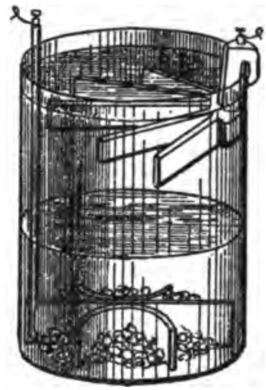


Fig. 431.

The *dry* cells, of which there are many sorts on the market, are adapted especially for intermittent service, such as ringing bells or operating signals. On short circuit they polarize after a time, though they partially recover if allowed to rest. The general principle of these cells is that of the *Leclanché* which has for a positive element a plate of carbon packed in manganese dioxide. The negative pole is of zinc and the electrolyte is an ammonium chloride solution. The chemical reactions may be expressed as follows, $Zn + 2NH_4Cl + 2MnO_2 = ZnCl_2 + 2NH_3 + Mn_2O_3 + H_2O$.

664. Energetics of a Voltaic Cell. The E. M. F. of a voltaic cell can in certain cases be calculated at once from the chemical reactions that take place. For E is the difference of potential of the plates of the cell when they are connected through a sufficiently great external resistance, and this equals the work done in transferring unit of electricity from the plate of lower potential to that of higher. Now let h be the decrease of chemical energy caused by the reactions that take place in the cell when one coulomb of electricity passes through it. If the electrical energy produced is equal to the decrease of chemical energy $E = h$. The energy given up by the chemical reactions can be calculated from the known heat that these reactions produce, when they do not take place in a voltaic cell. In the Daniell's cell, when one coulomb passes, the net chemical result is that 0.0003387 g. of Zn displaces 0.0003293 g.

of Cu from combination with SO_4. The total heat that would thus be given up can be readily calculated from thermochemical data (§ 267) and is found to be 0.260 calories or (multiplying by 4.2 the number of joules in a calorie) 1.09 joules. Now 1.09 is very closely the E. M. F. of a Daniell's cell. Hence we are entitled to conclude that, in this cell, the electrical energy produced is simply equal to the chemical energy expended.

If, however, we apply the same method to the simple cell of Volta, the result is different. Here when one coulomb passes 0.0003387 g. of Zn dissolves in H_2SO_4 and this produces 0.197 calorie or 0.83 joules. Hence if the electrical energy produced were equal to the chemical energy expended, the E. M. F. would be 0.83 volts. In reality the E. M. F. of the cell is 1.06 volts. The full explanation of this difference and of similar differences in other cells was first given by Helmholtz, who showed that, in general, part of the electrical energy comes from other than chemical sources, that, in fact, as the current passes heat is taken up by *reversible* thermal effects (§ 679) at the various contacts. This was experimentally verified by Jahn by placing a cell in a calorimeter. From the principles of thermodynamics Helmholtz deduced a complete formula for the E. M. F. of a cell. If c is the temperature coefficient of a cell, that is, the change in its E. M. F. when the temperature of the whole cell is changed by 1°, and if the cell is at $T°$ absolute,

$$E = h + cT$$

Since c can be found experimentally, E can be calculated and Helmholtz's formula is thus verified. In Daniell's cell c is extremely small, which accounts for the fact that the simpler formula $E = h$ gives the E. M. F. of a Daniell's cell.

665. Standard Cells. For many accurate electrical measurements a definitely known E. M. F. is absolutely necessary and several forms of *standard cell* have been devised to meet this requirement. In no case are they used to furnish current, but only to establish a potential difference that can be accurately known and relied upon at any definite known temperature. The Clark cell meets these requirements. As at present constructed it is con-

tained in a glass vessel of the shape shown in Fig. 432. At the bottom of the left hand part is placed the anode, consisting of a 10 per cent. zinc amalgam, above which are packed small crystals of $ZnSO_4$. In the other limb at the bottom is mercury over which is a layer of a paste formed of mercurous sulphate and a saturated zinc sulphate solution. A saturated zinc sulphate solution then fills the vessel to a point near the top, where it is sealed with a stopper. As the concentration of the zinc sulphate solution depends on the temperature, and for other reasons, the resulting E. M. F. is a function of the temperature. The following formula gives the E. M. F., E, at a temperature $t°$,

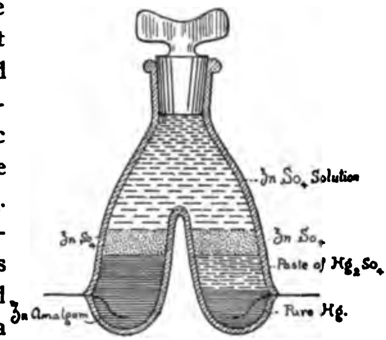


FIG. 432.

$$E_t = 1.433 - 116 \times 10^{-5}(t - 15) - 10^{-6}(t - 15)^2$$

Another standard cell, known as the cadmium cell, is constructed exactly like the Clark cell with the exception that cadmium is used as an electrode instead of the zinc, the electrolyte being a saturated solution of cadmium sulphate. The E. M. F. of this cell at $t°$ is given by the formula

$$E_t = 1.0184 - 3.8 \times 10^{-5}(t - 20) - 0.065 \times 10^{-6}(t - 20)^2$$

The cadmium cell excels the Clark cell in the fact that its temperature variation is considerably less.

In accordance with the International Electrical Congress, Chicago, 1893, and adopted by act of Congress in 1894, the legal definition of the volt in the United States is that unit of electromotive force which "shall be what is known as the international volt, which is the electromotive force that, steadily applied to a conductor whose resistance is one international ohm will produce a current of one international ampere, and is *practically equivalent*

to $\frac{1000}{1434}$ of the electromotive force between the poles or electrodes of the voltaic cell known as Clark's cell, at a temperature of 15° C. and prepared in the manner described in the standard specifications."

666. Various Arrangements of Cells. In the use of voltaic cells and other current generators two or more are often necessary to produce a desired result, and in joining them, one of two or three methods must be chosen. The two general ways are *in series* and *in multiple*. The former consists in arranging the cells with the negative pole of the first joined to the positive of the next, and the negative of the second to the positive of the third and so on till all are joined. The latter consists in joining all the positive poles together for one terminal and all the negative poles together for the other terminal. When cells are joined in series the E. M. F. is increased in proportion to the number of cells, but the internal resistance is also increased, so that on short circuit the current may be little greater than could be obtained from a single cell. When n cells are joined in multiple, however, the effect is as if a single cell were used with each plate n times as large. But in this case the resistance would be $1/n$th as great as that of a single cell. Therefore on short circuit the current would be nearly n times as great as from a single cell. If, however, the external resistance is necessarily a large part of the whole, then the maximum current is obtained by joining the cells in series. A combination of both methods of joining cells is often useful.

Under what circumstances a maximum current may be obtained with a given number of cells and a given external resistance may be found by an application of Ohm's law. Suppose we have n cells, each of E. M. F. e and resistance r, and suppose the external resistance to be R. Let the cells be joined, x in series, and suppose there be y such series joined in multiple Fig. 433. Then from what has been said the current will be given by the formula.

$$I = \frac{xe}{\frac{xr}{y} + R}$$

This may also be written

$$I = \frac{ne}{(\sqrt{xr} - \sqrt{yR})^2 + 4\sqrt{nyrR}}$$

Hence I is a maximum when $xr = yR$. But xr/y is the total internal resistance of the battery, while R is the external resistance. It follows, then, that with a *given number* of cells the maximum current is obtained when the cells are so joined, that the ratio of the internal resistance to the external resistance shall be as nearly as possible equal to unity. It is evident that in any given case the conditions for maximum current can be only approximately realized, since a cell cannot be subdivided.

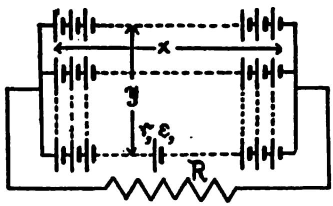


FIG. 433.

This is evidently not the arrangement that will result in the maximum efficiency. Since the energy expended in any part of the circuit is proportional to the resistance of that part, the energy wasted on the battery is evidently 50 per cent. of the whole, and the useful efficiency is only 50 per cent. Some other arrangement of the cells which would result in a less current would also result in a larger part of that current being usefully employed in the external circuit. The maximum efficiency will evidently result when the internal resistance is least.

ELECTROTHERMAL EFFECTS.

667. Joule's Law. If a quantity Q of electricity be moved from one point to another against a potential difference V, the potential energy produced will be QV; and if the quantities be expressed in C. G. S. electromagnetic units, the energy will be expressed in ergs. If now Q flow uniformly in time t from the point at the higher potential to that at the lower, the potential energy W will be expended on the conductor, and since $Q = It$, we shall have $W = ItV$. If I be expressed in amperes and V in volts, W will be expressed in joules. This follows from the fact that one ampere is 10^{-1} absolute units, and one volt is 10^8 absolute units; therefore their product, which is 10^7 absolute units of work per second must be in joules per second, as one joule is 10^7 ergs.

As $V = IR$ from Ohm's law, we may write

$$W = I^2Rt$$

Since then $W = VIt = I^2Rt$, we may state that the work done per second by one ampere between two points in a conductor

differing in potential by one volt, or between which there is one ohm resistance, is one joule. Because the relation expressed in the above equation was first established experimentally by Joule, it is usually known as *Joule's law.* The energy appears as heat in the conductor, and in order to reduce W, expressed in ergs, to calories, we must divide by the mechanical equivalent of heat, or by 4.19×10^7 (§ 316). In this case

$$H = 0.2387 \times 10^{-7} I^2Rt \text{ calories}$$

all units being C. G. S. electromagnetic units.

As in most cases the heat developed in a conductor is waste energy, it follows that for a given current it is desirable to have as little resistance as possible. From Joule's law we may write at once

$$A = W/t = I^2R$$

in which A is the power or activity required to maintain a current I through a resistance R. If I and R are expressed in C. G. S. units, A will be given in ergs per second. If however I be in amperes and R in ohms, then A will be in joules per second or watts.

668. The Electrical Measurement of the Mechanical Equivalent of Heat (see §§ 295–297). Evidently Joule's equation may be written,

$$HJ = I^2Rt$$

both sides being expressed in calories, and a very accurate method for determining the value of J has been developed, especially by Griffiths, by immersing a coil of platinum wire in a water calorimeter and passing a known current through the coil for a measured time t. If now H be measured by the usual calorimetrical methods and R be determined, all the quantities except J will be known and may be used in the formula to get J.

669. Electric Heating. If a sufficiently strong current be forced through a coil of wire, its temperature can be raised to any degree up to that at which the wire will be fused and the circuit broken. Various heating devices are constructed on this principle. The material should have a rather high specific resistance and a low specific heat. For domestic purposes it is desirable that the tem-

perature of the wires should not be high enough to be in danger of setting fire to ordinary objects. For increasing the heat it is desirable to call into service more wire rather than to raise the temperature.

670. Safety Fuses. The fact that the temperature may get so high as actually to melt a wire and thereby open a circuit, is made use of in practice, to protect circuits and instruments from injury by too strong a current. A short piece of fusible wire of proper dimensions is inserted somewhere in the circuit, often enclosed in a protecting tube. If by carelessness, ignorance, or by accident, that part of the circuit becomes overloaded, the wire will melt and the circuit will be opened. The device is called a *safety-fuse*.

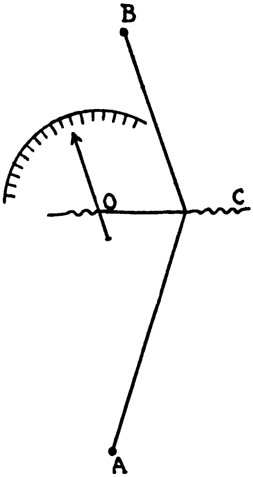


FIG. 434.

671. Hot-wire Measuring Instruments. The expansion of a conductor, owing to the heating effect of an electric current passed through it, is made use of in some instruments to measure current strength. Suppose, for example, between the points *A* and *B* in an instrument, Fig. 434, a fine wire is fastened of such a length as to equal *AB* when quite cold. At the middle of the wire *O* is fixed a light thread extending in both directions perpendicular to the wire. On one side it is drawn taut by a spring fixed at *C*. In the other direction the thread is passed a couple of times around a small drum, to which is attached an index needle moving over a graduated scale, and is then attached to another spring of somewhat less tension than the former. From the description the action is clear. If a current passes, the wire is heated and expands, the slack being taken up by the spring *OC*. The motion of *O* will cause the drum and needle to rotate, the amplitude of the motion depending on the rise in temperature caused by the current. If there were no losses this would depend on the square of current strength. In practice, however, the scale is empirically graduated by the comparison of the instrument with a standard. As in this instrument the resistance is often very high, owing to the fineness of the wire, it is usually designed as a voltmeter rather than as an ammeter (§ 641).

ELECTRIC LIGHTING.

672. The Incandescent Lamp. It has been seen that if the current be strong enough in a conductor, the latter gets hot and, if this heat be intense enough, the conductor becomes white hot and so is a source of light. The history of the development of this industry is long and interesting but would be out of place in a book of this character. The modern incandescent lamp consists of a fine conducting filament enclosed in a pear-shaped glass bulb exhausted of air, the terminals of the filament being led through the glass by means of two short platinum wires. For this purpose platinum must be used, because it has an expansion coefficient about the same as that of glass, and so, when the bulb becomes somewhat heated, as it does when in use, it neither leaks air nor is broken. Edison first devised this form of lamp in 1879 and used a filament made of bamboo and carbonized at a very high temperature. The bulb must be exhausted, or the filament would be consumed by igniting with the oxygen of the air. The earlier lamps were rather short lived because of the breaking of the filaments due to various causes. Very many materials besides carbon have been experimented with for filaments, the effort being to produce a lamp of *long life* and of *high efficiency*.

The efficiency of an electric lamp is defined as the ratio of the electrical power expended to the illumination in candle-power emitted. It is ordinarily expressed in watts per candle-power. The ordinary commercial unit is 16 candle-power and this lamp is used mostly on an E. M. F. of 110 volts taking about 0.5 ampere. As the power consumed is equal to EI, the ordinary lamp requires about 55 watts or about 3.5 watts per candle-power. As the light emitted by the carbon filament increases far more rapidly than does the power expended, the economy is rendered greater by increasing the current. But the high temperature resulting decreases the life of the lamp, so that electrical efficiency must be sacrificed to some extent to durability, and the net commercial efficiency is a maximum in any given type of lamp when the saving due to an increase in electrical efficiency is just balanced by the loss due to a decrease in durability. Filaments have been constructed of osmium, tungsten, tantalum, and some other ma-

terials, all giving a higher operating efficiency, some even 1 watt per candle-power, but so far the cost of production has been too high for general use. There is little doubt however that shortly we shall be able to produce a cheap durable lamp of an efficiency of 1 watt per candle-power or less.

The "glower" of the Nernst lamp contains salts of zirconium, yttrium, and erbium, which radiate highly when heated. It is a poor conductor when cold; hence a heater, consisting of platinum wire on kaolin, is placed beside (or around) the glower. The heater is automatically cut out when the glower is heated to the point of conduction. To prevent excessive rise of current when the glower conducts, a coil of fine iron wire (for "ballast") is put in series with the glower; the increase of resistance of the former compensates the decrease of resistance of the latter. The Nernst lamp is used mostly with alternating currents (in this country), but direct currents can be used. It can be made with an efficiency of 1.8 watts per candle-power.

673. The Electric Arc. As is well known there is another general method of lighting by electricity and it is by far the older of the two. Very early in the last century Davy produced a powerful light by placing in the circuit with an E. M. F. of 50 or 60 volts two carbon rods, which, after touching together, he separated a little. If the carbons are in a horizontal position the luminous stream is curved upward owing to the current of hot air, and the name of *electric arc* was given to this source of light. As a matter of fact, most of the light is emitted by the ends of the rods which became brightly incandescent. The positive rod is the hotter and is somewhat cupped out at the end, the negative one becoming pointed as the lamp burns. The positive rod burns away about twice as fast as the negative one. The lamp proper consists of the two carbons together with an appropriate mechanism for feeding them together at the proper rates. This regulating device is usually operated by the current and is then automatic. The efficiency of the arc lamp in furnishing light is much higher than that of the incandescent, yielding four or five times as much for the same expenditure of power. An ordinary lamp taking 800 to 1000 watts gives 1000 candle-power of illumination.

674. Other Forms of Arc Lamp. Other arc lamps, in which the light comes chiefly from the stream between the terminals are now in common use, and these lamps have a much higher efficiency than the carbon arc. When *magnetite* is used as the negative terminal and copper as the positive, the arc flame is of great brilliancy and whiteness, and an analysis of the light shows that it comes from luminous iron vapor. "Flaming arcs" are made by impregnating the positive terminal with salts of calcium. A very remarkable form of arc light of great length is the *mercury vapor lamp*, which consists of a long glass tube, exhausted to a high vacuum, and containing a pool of mercury as negative terminal, and a rod of iron or graphite as positive terminal. The tube is first tilted, so that a current passes through the liquid mercury; when the tube is righted, the current continues to flow through the mercury vapor. A continuous stream of vapor issues from a varying point on the surface of the pool and moves toward the positive terminal with a high velocity. This vapor is not at a high temperature, but it gives off a large amount of the green and yellow light characteristic of the spectrum of mercury. An alternating current, of ordinary voltage, will not serve for such a lamp (of the simple form here described); for the vapor stream ceases when the current is in such a direction that the mercury is the positive terminal, and it does not start up again when the current is reversed. But if the vapor stream from the mercury pool be sustained, by connecting a direct current to the mercury and to a third terminal in the tube, every alternate pulse of the alternating current will pass, that is, the alternating current will be *rectified* or changed into a direct current, consisting of pulses in one direction only. (By the use of devices, into which we cannot enter, the direct current may be dispensed with.) This is the principle of the *mercury arc rectifier* for alternating currents.

675. The Electric Furnace. As has been noted the temperature of the electric arc is the highest that can be produced and may reach the upper limit estimated if energy enough be supplied. For use in melting very refractory substances and for promoting chemical processes which require very high temperature, Moisson devised an electric furnace, consisting of a block of lime with a suitable cavity scooped out, into which the crucible containing the substance to be heated is placed. Through the opposite sides of the block a pair of thick carbon rods are passed into the cavity. A very strong current maintaining the arc between these terminals furnishes the heat.

676. Electric Welding. The method of electrically welding

together rods and wires was devised by Elihu Thomson who used the alternate current transformer (§ —) to produce currents of hundreds of amperes at an E. M. F. of only a few volts. If two metal rods or wires be forced together longitudinally, while such a current is passed through them, the junction becomes hot enough to melt the opposing layers and firmly weld the rods into a single piece. When this is completed the rod has become uniform and the heat is no longer localized.

THERMOELECTRIC EFFECTS.

677. The Thermoelectric Junction. The pages just preceding have demonstrated that the conversion of electric energy into heat, which is called the Joule effect, is easy and complete. We shall now consider the reverse process, which is most difficult, except in a very inefficient way. In 1821 Seebeck, while studying potential differences in general, discovered that, if a circuit be composed of copper and iron wires joined at their ends, there is a current in the circuit, unless the two junctions are at the same temperature, the current flowing from the hot junction to the cold in the iron. The potential difference causing the current depends upon the particular substances employed, upon the difference in temperatures of the two junctions, and also upon the mean temperature of these junctions.

Thus let Fig. 435 represent a circuit composed of a copper and an iron wire joined at the points A and B. If A be kept at 0°

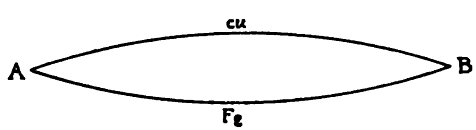


FIG. 435.

while B is gradually heated, a current will flow from copper to iron through B, increasing in strength as the temperature rises till a temperature of about 275° is reached. As this temperature, called the *neutral point,* is exceeded, the current will begin to decrease in strength till the temperature of B has reached about 550°; beyond this temperature, called the *temperature of inversion,*

the current is in the opposite direction. This phenomenon of inversion was discovered by Cumming in 1823. Fig. 436 illustrates these relations. As the resistance is nearly constant the currents are proportional to the varying E. M. F.'s measured in micro-volts, *i. e.*, 10^{-6} volts, which are plotted as ordinates, while temperatures are plotted as abscissas. Curve *a* illustrates the case just described, while *b* is the same when the junction *A* is kept at 100° instead of at 0°. Curve *c* is a zinc-iron diagram in which the neutral point is about 200°. It must be noticed that the neutral point depends only on the two metals concerned and is constant for these two, while the temperature of inversion is variable and is always as much above the neutral point as the temperature of the other junction is below it.

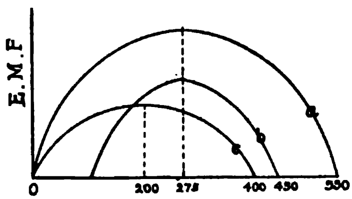


FIG. 436.

678. Thermoelectric Power. An examination of the curves in Fig. 436 shows them to be approximately parabolas. Since potential differences are ordinates and temperatures are abscissas the equation may be written

$$V = at + \frac{b}{2}t^2$$

where *a* and *b* are constants depending on the particular metals chosen. It is usual to choose some metal as a standard and to refer all others to this one. The one adopted is lead. Let the curve Fig. 436 represent the *V, t* relations in the case of a lead-iron circuit, one junction being at 0°. In this case it has been found experimentally that the neutral point is 357° and the temperature of inversion 714°. If the potential difference developed were proportional to the rise of temperature of the hot junction, the above curve would of course be a straight line. This however is not the case and the *rate* at which *V* varies with *t* is determined at any temperature by drawing a tangent to the curve at that point and measuring the tangent of the angle it makes with the axis of temperatures.

This equals dV/dt or from the equation of the curve

$$\frac{dV}{dt} = a + bt$$

This is known as the *thermoelectric power* of the two metals, and is seen to be a linear function of the temperature.

When dV/dt is zero V is a maximum and $t = t_n$ (the neutral point) $= -a/b$. Hence from the first equation $V_{max} = -a^2/2b$, and so, substituting the experimental values for t_n and V_{max}, we can determine the values of a and b for any particular pair of metals.

The following table gives an idea of the relation of several of the common metals to *lead* as determined by the values of a and b, and of the neutral points.

	Neutral point.	*a.*	*b.*
Iron	+ 357°	+ 17.34	− 0.0487
Tin	+ 45	− 0.43	+ 0.0055
Zinc	− 32	+ 2.34	+ 0.0240
Copper	− 68	+ 1.36	− 0.0095
Cadmium	− 69	+ 2.66	+ 0.0429
Aluminum	− 113	− 0.77	+ 0.0039
Silver	− 115	+ 2.14	+ 0.0150

679. The Peltier Effect. What may be deemed a converse of the Seebeck effect was discovered by Peltier in 1834, viz. that when an electric current is sent through a junction of two metals, the result is either an absorption or an evolution of heat according to the direction of the current. If the current passes in the same direction as that developed by heating the junction, then absorption of heat occurs at the junction, but if in the opposite direction, evolution of heat at the junction results. This is known as the *Peltier effect*. As this effect is never very intense, it is often masked by the Joule effect resulting from the resistance of the circuit, since the former is proportional to the first power only of the current, while the latter is proportional to the square of the current. A simple way of demonstrating the Peltier effect is to form a cross of bismuth and antimony rods joined in the middle, Fig. 437. If a difference of potential is maintained between *A* and *B*, by joining these points to the poles of a current generator of some sort, a current will flow, say, from *A* through the junction to *B* as indicated by the arrows. This is in a direction opposite to that which would result if the junction

were heated. The effect of such a current, according to the law stated, will be to heat the junction. This heat will in turn cause a current to flow from the bismuth to the antimony through the other ends of the cross which current will be indicated by a galvanometer connected between A' and B'. If the generator current be reversed, the junction will be cooled, and, if the whole be at 0° at first, a drop of water on the junction may be frozen by the cooling effect of the current. If a current be sent through a bar made up of three pieces in the order antimony-bismuth-anti-

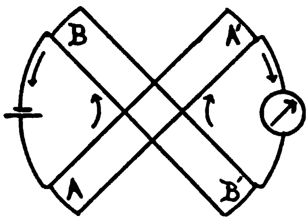


FIG. 437.

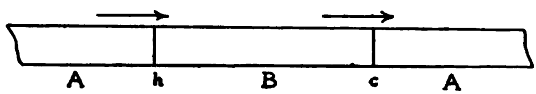


FIG. 438.

mony, Fig. 438, the heating and cooling effect will be produced simultaneously, heat at *h* where the current passes from antimony to bismuth, and cold at *c* where the current flows from the bismuth to the antimony.

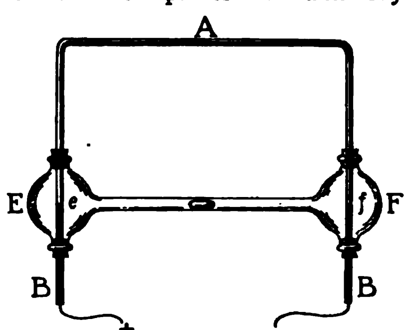


FIG. 439.

680. Measurement of Peltier Effect. Edlund devised an appliance for measuring this effect. A bismuth rod turned twice at right angles is terminated by an antimony piece at each end. Each junction is surrounded by an air-tight bulb filled with air, the bulbs being jointed across by a fine tube containing a short thread of mercury, the position of which is in the middle when the temperature is the same at the two junctions. If now a current be sent through the rods in the direction *eAf*, the right hand junction will get warm while the left hand one will become cooler then the rest of the circuit. The result is to push the mercury drop towards the left, because of the increase of air pressure on the right and the decrease of it on the left. The result of the Joule effect is the same

on both sides and so has no effect on the mercury, whatever motion there is being due entirely to the Peltier effect.

681. The Thomson Effect. Suppose a thermoelectric circuit as in Fig. 440. An iron wire I passes through the bottoms of two glass vessels and at both ends is joined to a copper wire c which completes the circuit above the vessels. Suppose water and ice are placed in the left hand vessel and water at a temperature t_1 is put in the other vessel. A current will now flow around the circuit in the direction indicated by the arrows, *i. e.*, through the hot junction from the copper to the iron. But, according to the Peltier law, there will be an absorption of heat at the right hand junction which will be supplied from the warm water, and an evolution of heat at the left hand junction which will melt some of the ice. We see, then, that the thermoelectric current is accompanied by a transfer of heat from a "source" to a "refrigerator," and if the amount of heat that disappears from the "source" is just equal to that which appears in the "refrigerator," added to what is developed in the circuit by the Joule effect, we have a striking analogy to a thermodynamic engine and the maintenance of the current is accounted for. If an electric motor were included in the circuit the analogy would be still more complete. Again if we arrange things, as is quite possible, so that the Joule effect is very small and so that no appreciable heat conduction occurs along the circuit, we shall approximate to a reversible heat engine. From the thermodynamic considerations, then, it follows that if t_0, the temperature of the cold junction, be kept constant, the E. M. F. of the circuit would be proportional to the difference in temperatures between the hot and cold junctions. But this is not consistent with the phenomenon of inversion discovered by Cumming. For this reason Kelvin was led to look for some other reversible thermal effect due to the current, besides the Peltier effect. The result was that he discovered that when a current is passing in a conductor, in which the temperature is different in different places, heat is liberated at a given point when the current is in one direction and absorbed if the current is in the opposite direction. The effect of the direction of the current depends on the substance. In the case of copper there is heat absorption when the current is from cold to hot in the wire, while the event is reversed when the current is reversed in direction. In any one case the reversing of the current reverses the thermo-effect. This phenomenon is known as the *Thomson effect*. It was shown by Le Roux in 1867 that the Thomson effect is practically zero in lead and, for this reason, lead is chosen as the metal of reference in the thermoelectric diagram.

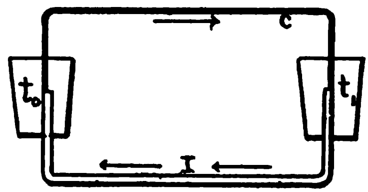


FIG. 440.

The net E. M. F. of a thermoelectric couple is due to the algebraic sum of the potential differences resulting from the Peltier effect and the Thomson effect.

682. The Thermopile. Many attempts have been made to use the thermoelectric principle in the design of apparatus for producing electrical energy directly from heat on a large enough scale to be of use.

Since, however, the E. M. F. of a bismuth-antimony pair when the junctions are at 0° and 100° is only about 0.01 of a volt, it would require more than 100 pairs to yield one volt. Such an arrangement, called a thermopile, Fig. 441, is made by joining the pairs in such a way as to bring one set of alternate junctions to one end of the bundle while the other set is at the other end. Thus it is easy to heat one set and to cool the other so that the various E. M. F.'s all act together in the same direction. But it has been found that the energy efficiency of the thermopile is extremely small, only a fraction of one per cent. of the energy of the heat applied being recovered in the current. The direct conversion of heat into electrical energy, either by this method or any other, has, up to the present time, proved to be of little value commercially because of the small efficiency. The thermopile has, however, a valuable place among the sensitive instruments for detecting and measuring small differences of temperature. For this purpose it is used with a sensitive galvanometer. The use of this instrument in the investigation of radiant energy has already been mentioned.

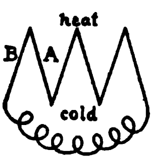


FIG. 441.

ELECTROMAGNETIC INDUCTION.

By A. P. Carman, Sc.D.,

Professor of Physics in the University of Illinois.

683. Induced Electric Currents. On November 24, 1831, Michael Faraday described to the Royal Society of London a series of experiments showing that electric currents can be produced in a closed conducting circuit, (*a*) by moving neighboring magnets; or (*b*) by changing the current in a neighboring electric circuit; or (*c*) by moving a neighboring electric circuit. An electric current thus produced is said to be *induced,* and the phenomenon is called *electromagnetic induction.* Few discoveries in science have had such important practical results as this discovery of Faraday's. Almost every modern industrial application of electricity depends upon electromagnetic induction.*

684. Faraday's Experiments. The experiments on induced currents made by Faraday were the following: (I.) A coil of wire *B* forms a closed circuit through a sensitive galvanometer *G* (Fig. 442). When the pole of a **magnet** is brought up to *B*, a momentary current is induced in *B*, and the galvanometer needle is deflected. When the magnet pole is removed, a momentary current is again induced, but in the opposite direction to that upon approach. The following facts may be noted: (*a*) The essential motion is relative, that is, moving the coil to or from the magnet produces the same effect

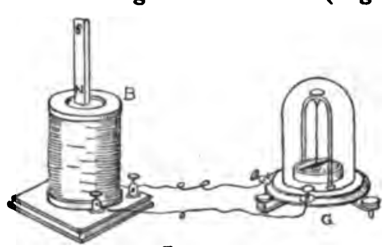


Fig. 442.

* Working at the same time, Joseph Henry discovered independently the fundamental facts of electromagnetic induction, and probably even anticipated Faraday in some cases. But Professor Henry worked under many disadvantages in the isolated town of Albany, New York, and his discoveries were not widely known at the time they were made.

as moving the magnet; (*b*) the current lasts only during the time of motion; when the magnet and the coil are relatively at rest, there is no induced current; (*c*) bringing up a *N* pole to a coil induces a current anti-clockwise as seen from the pole; that is, the induced current makes this face a *N* face (§ 632). Thus the approaching *N* pole is repelled by the magnetic action of the induced current. Removing the *N* pole induces a clockwise current, that is, makes the coil face a *S* face. Thus the *N* pole is attracted as it is removed. An approaching *S* pole induces a current in the same direction as a receding *N* pole, and *vice versa* (Fig. 443). Or in general, **the magnetic action of the induced current opposes the motion of the magnet.** This is evidently a case of action and reaction. If the approaching magnet were attracted by the induced current, it would require no work to bring the magnet up, and we would get an electric current, which represents energy, without the expenditure of work. This would be contrary to the principle of the conservation of energy (§ 318).

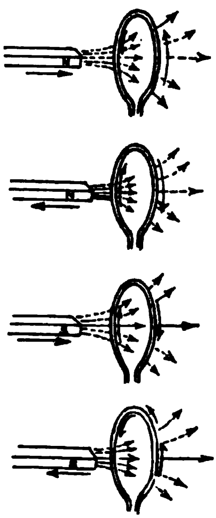


FIG. 443.

We can now describe the above experiment in the convenient terms of the magnetic field and the magnetic lines of force, as conceived by Faraday (§ 559). The magnet is surrounded by a magnetic field, and the lines of force emerge from the *N* pole, and enter at the *S* pole. The motion of the magnet thus changes the number of lines of force included by the coil. The experiment thus shows that, (*a*) a change of the number of lines of magnetic force included by a circuit induces a current in the circuit; (*b*) the current induced is proportional to the rate of change of included lines of force; and (*c*) the magnetic lines from the induced current increase as the magnetic lines from the magnet decrease through the circuit, and vice versa. The positive direction of change of lines is to be reckoned in the same direction for both magnet and current. Faraday's other experiments are now easily described.

(II.) Substitute for the magnet *NS*, a **coil carrying an electric current.** *A* is thus surrounded by a magnetic field (§ 622), and moving *A* in front of *B*, changes the number of magnetic lines included by *B*, and thus induces an electric current in *B* during the time of motion. When *A* is approaching *B*, the opposing faces of the two coils are either both *N* or both *S*; and for this case the induced current in *B* must be ***inverse*** in direction to the current in *A*. Similarly it is seen that upon the receding of *A*, the induced current is ***direct*** to that in *A*.

The coil *A* carrying the original or inducing current, is called the primary coil (*Pr*), and its current the primary current. The coil *B* is called the secondary coil (*Sc*), and the induced current, the secondary current. A current in the same direction as the primary current is called ***direct***, and a current in the opposite direction is called ***inverse***.

(III.) With **two coils** *A* and *B* as before, we can change the number of lines of force through *B*, by changing the current in *A* (Fig. 444). Thus we find that, ***making*** or ***increasing*** the *Pr* current induces a momentary inverse current *Sc* in *B*; and that, ***breaking*** or ***decreasing*** the *Pr* current induces a ***direct*** current in the secondary circuit.

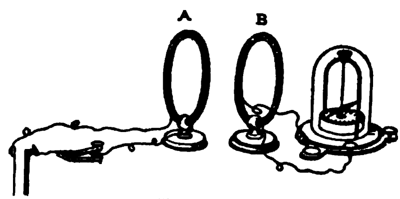


FIG. 444.

The coils *A* and *B* may be placed anywhere, provided that the magnetic lines from *A* pass through *B*. Coils wound over or alongside of each other on a cylinder are common arrangements for obtaining maximum induced currents (Fig. 446). Two straight parallel wires may similarly form primary and secondary circuits. This is often the case in telephone circuits, and is the cause of the "cross talking" in the lines.

(IV.) If *A* and *B*, one or both, have an **iron core**, the induced currents are greater, but in the same direction as without the iron cores. This is easily explained in terms of lines of force. Iron has a greater magnetic permeability (§ 565) than air, so that a given change in the primary current produces a greater change in the magnetic flux through the secondary circuit, and thus causes a

greater induced current. Figs. 445 and 446 show common arrangements of the primary and secondary coils on iron cores. The arrangement shown in Fig. 446 is one Faraday used in his earliest experiments.

685. Lenz's Law. In 1834, Lenz stated the following important relation between the induced current and the motion of the electrical circuit or the magnet causing the induction: *The induced current is in such a direction as to oppose by its electromagnetic action the motion of the magnet or the coil which produces the induction.* We have already seen that this holds in the case of the motion of a magnet, and it is easily seen that it also holds for two coils moving relatively to each other.

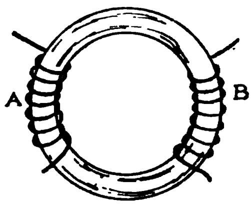


FIG. 445.

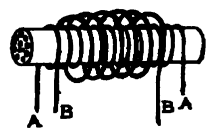


FIG. 446.

Lenz's law can be extended to the case of a current induced by the variation of a primary current, but the reactions are then purely electromagnetic. When the current is induced in a secondary coil B by making or breaking the current in a primary coil A, we have a reaction of B on the current in A. That is the electromagnetic induction is mutual. We thus have the following series of actions and reactions: Starting a current in A produces magnetic lines through A, and part of them pass through B. There is thus an inverse current induced in B. But this induced current started in B produces lines which are opposite to those produced by the primary current in A. The effect of the induced current is thus to oppose and retard the building up of the magnetic field through the coils.

Upon breaking the primary current, the current induced in B is direct; that is, this induced current produces magnetic lines which are in the same direction as those produced by A. The effect of the induced current in B is thus to maintain the field, that is, to delay the decrease of the number of lines of force through the coils.

It can thus be seen that in general, the magnetic action of the induced current opposes the magnetic action of the inducing current.

The study of the actions and reactions of the primary and secondary currents, with their energy relations, is very important in

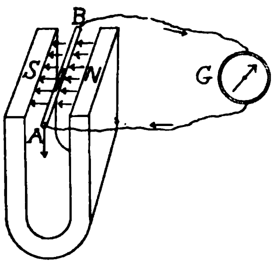


FIG. 447a.

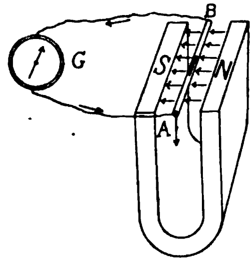


FIG. 447b.

understanding the complete theory of the induction coil (§ 695), and of the alternating current transformer (§ 703).

686. Induction by Cutting Lines of Force. One of Faraday's early observations was that "single wires, approximated in certain directions towards the magnetic pole [of a large electromagnet], had currents induced in them." It is often convenient to consider the induction as due to the motion of a single wire across lines of force, or as we often say, due to "cutting lines of force." Thus suppose the wire AB moved across the field between the poles N

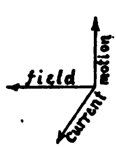


FIG. 448a.

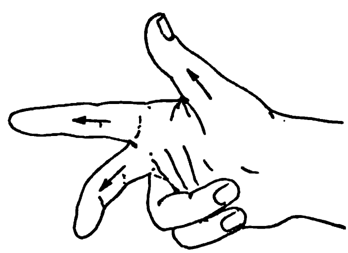


FIG. 448b.

and S (*e. g.*, of a U-shaped magnet). In case (a) the wire forms part of a complete circuit ABG. In moving AB down, the number of lines through the circuit, is increased, and an anti-clockwise

current (seen from below) is induced, that is from A to B in the part AB of the circuit. In case (b), the wire forms part of the circuit ABG, and the motion downwards induces a current in the circuit, so that the current flows from A to B. (Current clockwise, seen from below, as shown.) When the motion is upward, the current is evidently from B to A. The three directions, of magnetic field, motion, and induced current, are thus mutually at right angles, as indicated in the rectangular axes of Fig. 434*a*. Professor J. A. Fleming has given a convenient rule for remembering these relative directions. Holding the *right* hand as indicated in Fig. 448*b*, with the thumb, the forefinger, and the center finger, making right angles with each other, then if the forefinger is held in the direction of the magnetic field, and the thumb in the direction of the motion, the center finger will indicate the direction of the current.

687. Numerical Calculation of Induced E. M. F. By Ohm's law, the induced current varies directly as the induced electromotive force and inversely as the resistance of the curcuit. The resistance is a constant for the circuit (§ 625). Experiments show that the electromotive force induced in a circuit is proportional to the rate of variation of lines of force through the circuit, that is,

$$E = -K\frac{N_2 - N_1}{t}$$

where t is the time in seconds, and N_1 and N_2 are the number of lines at the beginning, and at the end of the time interval t. If the variation of N is uniform during the time t, then the E. M. F. induced is constant. When the variation is not uniform the E. M. F. at any instant is given by the differential coefficient, that is, $E = -K dN/dt$. In these expressions K is a constant. If E, N, and t are expressed in C. G. S. units, then it can be shown that K becomes unity. To express E in volts, we divide by 10^8 (§ 624), that is,

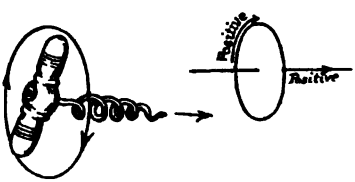


FIG. 449.

$$E(\text{volts}) = -\frac{N_2 - N_1}{10^8 t} = -\frac{1}{10^8}\frac{dN}{dt}$$

The negative sign is explained as follows: The positive direction of the lines of force is taken as that of the advance or thrust of a right-handed screw, Fig. 449, where the rotation of the same gives the positive direction

of the E. M. F., or current. Thus a positive increase of N corresponds to a negative induced E. M. F.

688. A Second Numerical Statement of Induced E. M. F. It is often convenient to calculate the induced E. M. F. in terms of the number of lines of force cut by a conductor per second. Let $H =$ the strength of the magnetic field, $=$ the number of lines of force per square centimeter section of the field (§ 559), and let $l =$ the length of the conductor in centimeters, and $v =$ the velocity in centimeters per second. If the motion is perpendicular to the lines of force and also to the length direction of the conductor, then the number of lines cut per second is lvH, or the induced E. M. F. is

$$E = lvH, \text{ or } E(\text{volts}) = \frac{lvH}{10^8}$$

In case the velocity v, and the conductor length are not at right angles to the field H, their components at right angles to H are to be taken. The equation for E has the same form as above, but l and v are components of l' and v', the actual length and velocity.

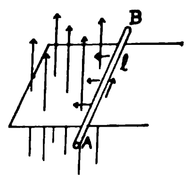


FIG. 449a.

689. Calculation for Current and Electric Quantity. In the above section (§ 687) it has been shown that the induced E. M. F.

$$E = -\frac{N_2 - N_1}{t} = -\frac{dN}{dt}$$

The current is then

$$I = \frac{E}{R} = -\frac{N_2 - N_1}{Rt} = -\frac{dN}{Rdt}$$

Thus we have $It = -(N_2 - N_1)/R$ and $Idt = -dN/R$. But $It = Q$, the total flow of electric quantity in the time t, and $Idt = dQ$, the electric quantity in the time dt. Thus the total quantity of electricity induced is

$$Q = -\frac{N_2 - N_1}{R} = -\int \frac{dN}{R}$$

This is the quantity measured by the throw of a ballistic galvanometer, when the time of induction is a small fraction of the period of the galvanometer needle (§ 640).

690. Faraday's Disk Dynamo. One of Faraday's earliest experiments in electromagnetic induction was to rotate a copper disk between the poles of a magnet, the plane of the disk being perpendicular to the field (Fig. 450). A galvanometer circuit was completed by wires sliding on the axle and on the circumference of the disk, and a current was shown by the deflection of the

galvanometer during the rotation of the disk. In this machine each radius of the disk cuts the lines of the field at the rate of $\pi r^2 nH$ per second, where πr^2 is the area of the disk, H is the

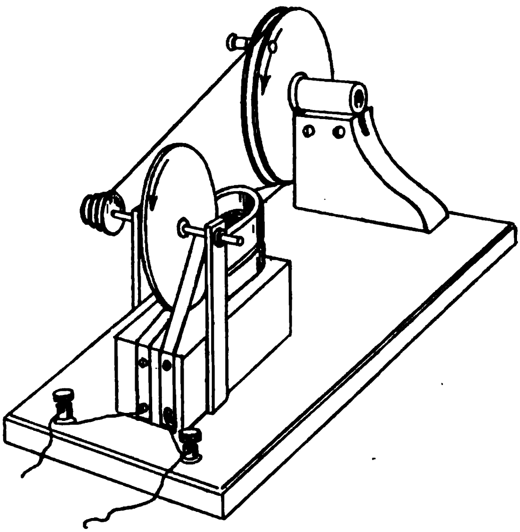


FIG. 450.

strength of the field assumed uniform, and n is the number of revolutions per second of the disk. Thus

$$E\,(\text{volts}) = \frac{\pi r^2 nH}{10^8}$$

This arrangement made the first dynamo-electric machine. Forbes and others have attempted to use this as a model for commercial electric generators, but the E. M. F., with any practical diameters and speeds is too small for industrial uses.

691. Foucault or Eddy Currents. It was observed a number of years before Faraday's discovery of induced currents, that a vibrating magnetic needle quickly came to rest when near or over a copper plate. Arago had in 1824 also shown that a magnetic needle suspended over a rotating copper disk, rotates with the disk (Fig. 451). Both the damping of the needle, and Arago's disk experiment were explained by Faraday as phenomena of

electro-magnetic induction. The relative motion of the magnet and the disk induces an E. M. F. in the metal disk. The current thus generated circulates in the disk, producing a magnetic action, which by Lenz's law tends to hold the magnet at rest relative to the disk or plate.

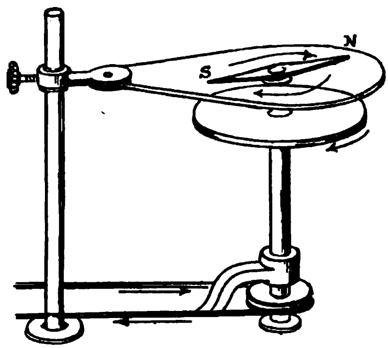


FIG. 451.

Electric currents, thus induced and circulating in a metal mass, are called eddy currents or Foucault currents. The energy of such currents is dissipated in heat. The iron cores of armatures of dynamo machines and transformers are always laminated so as to offer infinite resistance to the formation of such currents, and thus to stop the heat losses.

An interesting example of damping by eddy currents is shown in Fig. 452. The pendulum with its copper plate swings freely if the electromagnet is not excited, but is damped immediately when the magnetic field is made.

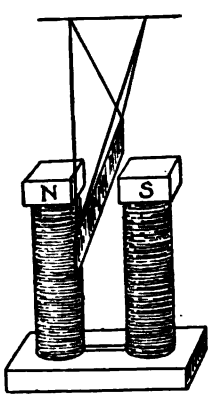


FIG. 452.

This damping action of eddy currents is often taken advantage of in D'Arsonval or movable coil galvanometers (§ 639), to bring the moving coil to rest quickly. The coil is wound on a closed copper frame in which the eddy currents are generated during the vibration. The coil itself is damped in the same way on closed circuit. The magnetic needles of galvanometers are also often damped by suspending them in openings in copper blocks.

692. Self Induction. Several persons seem to have observed independently about 1832, that there is a bright spark when a current is broken in a circuit containing an electro-magnet. On making a current in the same circuit, there is no such spark,

This extra current at break was noticed by persons receiving a shock upon breaking the circuit of an electromagnet, if they had the terminals in their hands at the time of break. In investigating this, Faraday observed the following facts. Upon breaking the circuit of a helix without an iron core, a similar bright spark is obtained, only less than in the case of an electromagnet. Likewise, when the current in a long straight wire was broken, a spark occurred, only less bright than in the case of the helix. Upon breaking a current in a short wire there was practically no spark. Also in the case of a long wire doubled back on itself, there was no extra spark at break. "The first thought that arises in the mind," wrote Faraday, "is that the electricity circulates with something like momentum, or inertia in the wire, and that thus, a long wire produces effects at the instant the current is stopped, which a short wire cannot produce. Such an explanation is however at once set aside by the fact that the same length of wire produces the same effects in very different degrees, according as it is simply extended, or made into a helix, or forms the circuit of an electromagnet." Faraday then proceeded to show that this extra current was due to electromagnetic induction, from the varying current acting on its own circuit. This phenomenon of a current inducing an extra electromotive force in its own circuit, is called *self-induction*. A circuit includes in general lines of force due to its own current. Breaking the current thus removes the lines of force, or has the same effect as removing a magnet. Suppose the current flows clockwise in a circular circuit; breaking the current then removes positive lines, or is the same as removing a north pole. But this induces a clockwise E. M. F., that is a direct current; this adds itself to the current being broken, and thus causes the bright spark of break. The effect of an iron core is to increase the magnetic induction, and thus to increase the E. M. F. of self-induction at break. Self-induction thus prolongs the current at break, or acts to retard a decrease of the current. When the circuit is wound back on itself, so that it includes no lines of force, there can be no change of lines of force, and hence no self-induction. Such circuits are said to be non-inductive or inductionless. The wire in resistance boxes

is wound doubled from its middle point as is shown in Fig. 453.

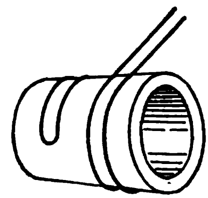


FIG. 453.

When a current is made or increased in an inductive circuit, such as a helix, magnetic lines are put through the circuit. Thus positive lines enter at the face in which the current is clockwise (S face), and this is equivalent to bringing up a N pole. But this induces an anti-clockwise E. M. F., that is an E. M. F. inverse to the starting current. That is, the building up of a current in a coil, is accompanied by an induced inverse E. M. F. at the time of the current increase. Here again the self-induction opposes and delays the current changes. Fig. 454 shows the growth and dying away of currents in an inductive circuit as observed in an oscillograph (§ 707). Helmholtz deduced in 1851 an equation showing the law of the growth of currents in inductive circuits and these oscillograph curves confirm the Helmholtz equation.*

FIG. 454.

693. Coefficient of Self-induction or Inductance. The E. M. F. of self-induction in a circuit thus depends upon the change in the number of lines of force through the circuit, caused by the variation of the current. The number of lines evidently depends upon

* Helmholtz's equation is

$$I = \frac{E}{R}\left(1 - \epsilon^{-\frac{R}{L}t}\right)$$

where I, E, R, L and t represent the quantities indicated in this and the next section, and ϵ is the base of the Naperian logarithms. This equation is deduced as follows: The E. M. F. in the circuit at any instant is equal to the impressed E. M. F. less the counter E. M. F. of self-induction, or $e = E - L\, dI/dt$. Hence the current is

$$I = \frac{E}{R} - \frac{L}{R}\frac{dI}{dt}$$

By integration of this differential equation we get the equation of Helmholtz,

$$I = \frac{E}{R}\left(1 - \epsilon^{-\frac{R}{L}t}\right)$$

(*a*) the current I, and (*b*) upon the dimensions of the circuit, and (*c*) the presence of magnetic substances, such as an iron core. For a circuit without iron, N the number of lines included by the circuit varies directly as I, or $N = LI$, where L is the *coefficient of self-induction* or the *inductance* of the circuit. Thus the inductance of a circuit is numerically equal to the increase in the number of magnetic lines included by the circuit for unit increase of the current. For a circuit with an iron core, this increase of magnetic lines per unit current is not constant, because the magnetic permeability of iron varies with the magnetizing field (§ 566). Hence L the inductance of a circuit with an iron core is a variable depending upon the magnetic curve of the iron.

We can also express the inductance of a circuit in terms of the E. M. F. induced for unit rate of change of the current in the circuit. This is shown as follows: The number of magnetic lines through a circuit is $N = LI$; hence the induced E. M. F. is $E = -dN/dt = -L(dI/dt)$. If the rate of change of the current is unity, that is if $dI/dt = 1$, then $E = L$. We can thus define unit inductance, as the inductance of a circuit in which unit E. M. F. is induced by unit change of current per second in the circuit. The practical unit of inductance is the *henry* and is equal to 10^9 C. G. S. units of inductance. The henry can be defined as the inductance of a circuit, in which a change of one ampere per second produces an induced E. M. F. of one volt. Standards of inductance are used in the shape of coils wound on marble or other non-magnetic and permanent cores. These are graduated in multiples or submultiples of the henry. A variable standard of inductance can be made by two coils joined in series and arranged so that they can be rotated in reference to each other, and thus change the total lines included. Such an inductance standard is illustrated in Fig. 455.

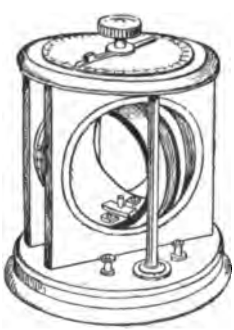


FIG. 455.

694. Experiments on Self-Induction. To demonstrate extra-currents

due to self-induction, Faraday made the following experiment: In a circuit containing a large helix or electromagnet M, there is a galvanometer G, the galvanometer being in parallel with the helix L (Fig. 456). The current from the battery B is made or broken by the key K. In the steady condition, the current divides between the helix and the galvanometer, and there is a steady deflection of the galvanometer needle, say of n degrees to the *right*. A stop is placed so that the needle can not deflect to the right. Upon breaking the current at K, there is a throw of the galvanometer to the *left*, due to the extra-current of break flowing back through the galvanometer circuit. Evidently the extra-current in M is in the same direction as the current being broken.

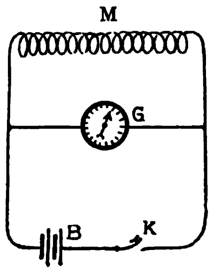


FIG. 456.

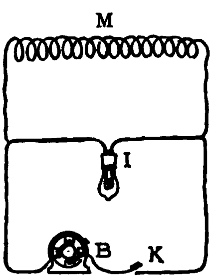


FIG. 457.

A striking variation of the above experiment is to put an incandescent lamp in parallel with an electromagnet, passing just enough current to bring the lamp to a red glow (Fig. 457). Upon breaking the current, the lamp flashes brightly for an instant, due to the E. M. F. of self-induction at break.

The best way of showing the effects of self-induction is by the use of the Wheatstone bridge, as described by Maxwell. In the bridge arrangement, Fig. 458, the resistances R_1, R_2, and R_3 are wound non-inductively (§ 692), and M is the coil or helix of an electromagnet. The resistances are arranged so that $R_1 : R_2 = R_3 : R_M$. There is, accordingly, no deflection of the galvanometer G when the current is in a steady state. But upon making the current by closing the key A, there is a momentary throw of the galvanometer needle, the deflection of the needle being again zero when the current reaches a steady state. Upon breaking the current, there is a throw of the needle in the opposite direction to that at make. If all four of the resistances are non-inductive, there are no such momentary throws of the galvanometer needle at make and break. Suppose the current enters at C; then the current reaches its full value in CR_2b sooner than in CLd, so that there will be a deflection of the galvanometer showing a momentary current from b to d. Upon break the momentary flow will be from d to b.

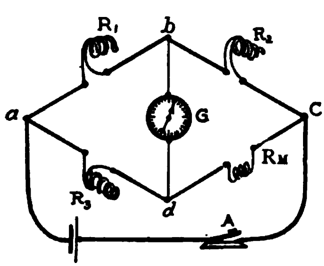


FIG. 458.

The explanation of this is evident from § 692. The above method of showing the extra-currents has been developed by Maxwell and others into a method of measuring the coefficient of self-induction. For these methods the reader must refer to the laboratory manuals.

695. The Induction Coil. The induction coil is a piece of apparatus for producing pulsating currents or discharges of high E. M. F. in a secondary circuit, by making and breaking a current in a primary circuit. The current in the primary circuit may be from a battery with only a few volts E. M. F. In Fig. 459, we have a diagram showing the parts and arrangement of the ordinary induction coil. The primary circuit *Pr* consists of (*a*) a solenoidal coil *P* with a bundle of soft iron wires as core; (*b*) an interrupter *K* for making and breaking the primary current. When the interrupter is mechanical, as shown in the figure, there is a condenser joined across the gap to lessen the extra-spark of break, and thus cause a quicker break of the current; (*c*) the secondary circuit *Sc* consisting of a solenoidal coil *S* surrounding the coil *P*, and a spark gap *D*. The secondary coil is wound with many turns of fine wire. To increase the insulation, this coil is also wound in disk sections. The primary coil is wound with a comparatively few turns of much larger wire.

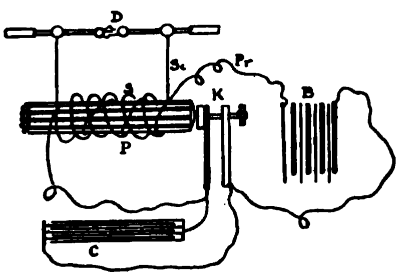


FIG. 459.

Making the primary current produces magnetic lines which thread through the secondary. These lines are removed upon breaking the primary current. Thus there is induced in the coil *S* an inverse E. M. F. at make, and a direct E. M. F. at break of the primary current. The break in most coils is much quicker than the make, and thus the direct induced E. M. F. in *Sc* is so much greater than the inverse induced E. M. F., that the discharge effects are mostly uni-directional. The reason for this is that the growth of the primary current at make is retarded by the inductance of the circuit (§ 692), while with a good interrupter and proper condenser, the break can be made very sharp. The greater the number of turns of the secondary coil, the greater

the induced E. M. F. The resistance of the coil is of course high, and consequently the current small.

In small induction coils the Wagner hammer is the common form of interrupter. This is shown at *K* in Fig. 459, and its action can be easily seen. In large coils, a form often used consists of a brush sliding on a revolving commutator driven by an electric motor. The electrolytic interrupter of Wehnelt is also frequently used (see Fig. 460). *P* is a platinum wire in a solution of sulphuric acid, *L* is a lead plate. Only the point of the wire is exposed to the acid. When *P* is made the anode, and *L* the kathode, gas forms at *P*, interrupts the current, and escapes in bubbles, and thus the current is again made.

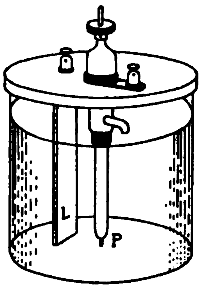


FIG. 460.

696. The Tesla Induction Coil. To obtain currents of very high frequencies and high electromotive forces, Tesla used a form of induction coil in which the oscillatory discharge of a Leyden jar (§ 722) is used as interrupter. The terminals of the secondary of an induction coil *I* (Fig. 461), are connected, one to the inner coating, and one to the outer coating of an insulated Leyden jar *C*. The circuit is completed through the primary winding of the Tesla coil, and the discharge balls. The primary of the Tesla coil *A*, consists of a half dozen turns of wires on a non-magnetic core. The coils *A* and *B* are separated by air or oil as insulation. The alternations at *D* from the Leyden jar may have a frequency of several millions per second (§ 722). Hence the currents induced in *B* are not only of high E. M. F. but also of very high frequency.

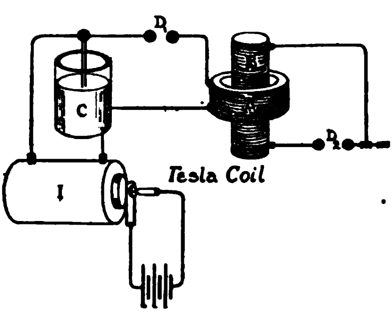


FIG. 461.

697. Electromotive Force in a Coil Rotating in a Magnetic Field.

In the earth inductor and in many dynamo-electric machines, electric currents are produced by rotating coils of wire in a magnetic field. We take the simple case of a rectangular coil in a uniform magnetic field. The coil $ABCD$ is rotated n times per second about an axis OO', which bisects the coil and is perpendicular to the field (Fig. 462). Let $l = AB = DC$, and $r = AO$

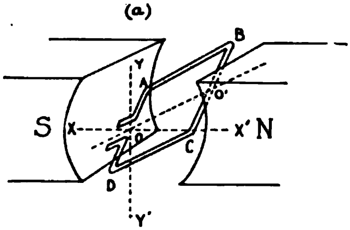


FIG. 462a.

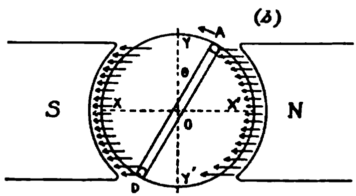


FIG. 462b.

$= DO$. The direction of rotation looking on the end AOD is anti-clockwise. Let XOX' and YOY' represent the planes through O, respectively parallel and at right angles to the magnetic field. The side AB is evidently cutting lines of force most rapidly at X and X', where it is moving at right angles to the lines; while at Y and Y', it is moving parallel to the field, so it is cutting no lines, and hence has no E. M. F. induced in it at this instant. The E. M. F. induced in AB is a maximum at X, and zero at Y, a negative maximum at X', and zero again at Y'. Thus the E. M. F. in AB is from A to B while the coil is moving from Y through X to Y'; it is from B to A while moving from Y' through X' to Y. (These directions follow in accordance with Fleming's rule, § 686.) The E. M. F. induced in the opposite side DC evidently adds itself to that in AB to produce a single alternating current in the circuit $ABCD$. The ends BC and DA do not cut lines, and hence have no E. M. F. induced in them. There is thus induced in the coil $ABCD$ an E. M. F. which reverses its direction once in every revolution. It can be shown that this E. M. F. at any instant is proportional to the sine of the angle θ which the plane of the coil makes at that instant with the plane perpendicular to the field. Thus in Fig. 462b, the E. M. F. is equal to $e = E \sin \theta$, where E is the maximum E. M. F. induced, and θ is the angle of the coil with the plane YOY'.

698. Formula for the E. M. F. in a Rotating Coil. Let $V=$ the uniform tangential speed of AB (and CD). At the instant when the angle between the coil and the plane perpendicular to the field is θ, this velocity is represented by AR, Fig. 463. The velocity component at right angles to the field is $SR=V \sin \theta$. Let $H=$ the strength of the field (= the number of magnetic lines per square centimeter); then $VlH \sin \theta$ is the number of magnetic lines cut by AB $(=l)$ per second. Hence the E. M. F. induced in AB and CD is $e=2VlH \sin \theta=E \sin \theta$. Here $E=2VlH=$ E. M. F. when the coil is passing through the points X and X' where it is cutting the lines at the maximum rate, or when $\theta=90°$ or $=270°$. The curve, Fig. 464, represents this E. M. F. during a single rotation. The ordinates are proportional to the E. M. F. and the abscissas are proportional to the angles. Since the rotation is uniform, the abscissas are

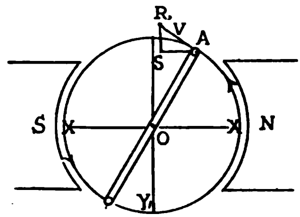


FIG. 463.

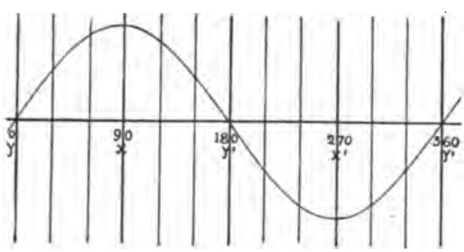


FIG. 464.

also proportional to the time. Ordinates above the line represent electromotive forces in one direction, and ordinates below represent electromotive forces in the reverse direction. That is, a coil rotating uniformly in a uniform magnetic field, has induced in it an alternating E. M. F. which varies as the sine of an angle, and is represented by a sinusoidal curve.

699. The Earth Inductor. The earth inductor is a coil (Fig. 465), usually of a large number of windings, which is mounted so that is can be rotated about either a horizontal or a vertical axis. Suppose the axis vertical and the plane of the coil at right angles

to the magnetic meridian. By revolving the coil through 180°, the magnetic flux is taken out of one face and put in at the other face. Let $H =$ the horizontal intensity of the earth's magnetic field, $S =$ the area of the coil face, and $R =$ the resistance of the circuit. Then the quantity of electricity flowing in the coil during a rotation of 180°, is $q = (2HS)/R$ (§ 689). This can be measured by the throw of a ballistic galvanometer. Similarly when the axis is placed horizontally, we get $q' = (2VS)/R$, where V is the vertical component of the earth's field. We thus get $(q'/q) = V/H = \tan \phi$, where ϕ is the dip or inclination (§ 571).

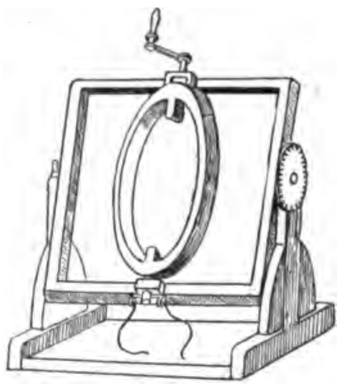


FIG. 465.

700. Simple Alternating Current Dynamo. In Fig. 466 we have a coil revolving in the field between the poles of a magnet. The ends of the coil are connected with the insulated metal rings N and M, on the shaft OO'. Metal springs or "brushes" rest or slide on these collector rings. Thus the current induced in the coil $ABCD$ flows through the external circuit. Such a machine forms a simple alternating current dynamo. By winding the coil on an iron cylinder, the intensity of the magnetic field is increased, and thus a greater E. M. F. is induced. But this iron core being itself a conductor will have eddy currents induced in it, unless it is laminated, so as to make the resistance infinitely great in the direction of the induced E. M. F. Hence the core is built up of insulated iron disks as represented in Fig. 467. Commercial alternating current dynamos are always multipolar. One of the reasons for this is that for practical electric

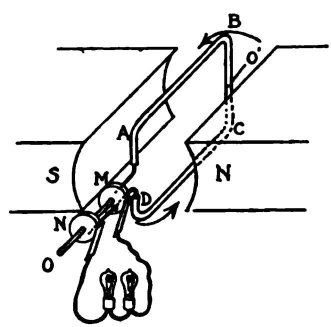


FIG. 466.

lighting and power transmission, frequencies of from 25 to 125 alternations per second are desirable. A common frequency for most purposes is now 60 alternations per second. To get such

FIG. 467.

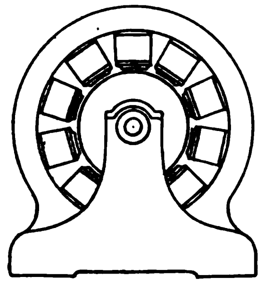


FIG. 468.

frequencies with safe speeds it is necessary to have field magnets with multiple poles. Fig. 468 shows a multipolar alternating current dynamo.

701. Simple Direct Current Dynamo. To obtain a current constant in direction, a commutator is used instead of collector rings. Fig. 469 represents a two-part commutator. This consists of a copper ring, which has been cut into half rings. These half rings are insulated and form the ends of the coil *ABCD*. The brushes *R* and *S* are set 180° apart, so that one rests on one-half of the commutator, while the other rests on the opposite half. These brushes thus make the connections for the current with the external circuit. The brushes are set so that the connections with the external circuit are reversed, just at the instant in which the current in the rotating coil is reversed (that is, approximately as the coil

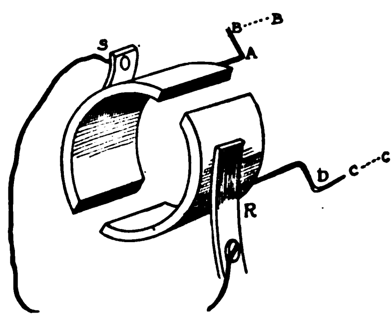


FIG. 469.

passes through the plane perpendicular to the field). The current in the external circuit is thus uni-directional, and varies

as represented in Fig. 470. The above is a simple direct current (D. C.) dynamo with a two-part commutator. In Fig. 471 we have two coils at right angles to each other, and so joined to a

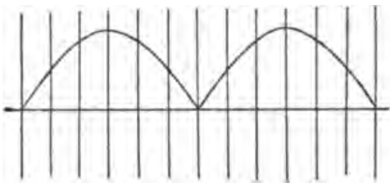


FIG. 470.

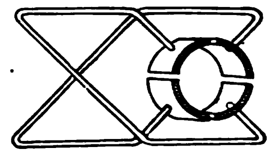


FIG. 471.

four-part commutator as to give in the external circuit, the current represented in Fig. 472. This is the sum of two pulsating currents, which differ in phase by 90°. It is seen that the per cent. of variation is much less than in the current from a dynamo with a two-part commutator. In modern direct current dynamos there are often hundreds of coils, joined to a commutator of many sections, and the resulting current is practically constant. The inductance of the circuit also operates to lessen the variations of the current in these machines.

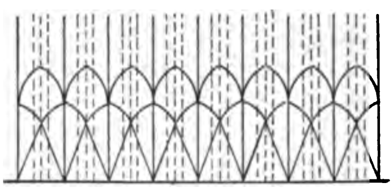


FIG. 472.

702. Dynamo-electric Machines. The parts of a dynamo-electric machine, or a dynamo, are (1) the field magnets for producing the magnetic field; (2) the armature, or the coils in which the currents are induced. The armature coils are nearly always on a laminated iron core, and are supplied with slip rings or a commutator, and brushes to make connection with the external circuit. In the simple forms of dynamos described in previous sections, the armature revolves, and the field magnets are stationary. In large dynamos for high electromotive forces, the armature is often made the stationary part and the field magnets revolve. In one type of A. C. dynamos, the revolving parts are iron masses which change the magnetic flux through the armature coils, the magnet coils being also stationary. This type of dynamo is called the "inductor" form.

Dynamos are direct current (D. C.), or alternating current (A. C.), according to the character of the E. M. F. at the terminals of the machine. The field magnets may be bi-polar or multi-polar. Several common forms of field magnets are shown in Fig. 473. In some of the early dynamos, permanent magnets

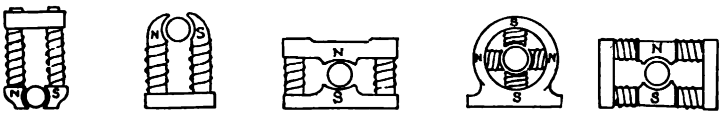


FIG. 473.

were used for field magnets. Such machines were called magneto-electric machines or magnetos. The small machines often used in telephone call boxes are magnetos. But the field magnets of all modern power dynamos are electromagnets.

Fig. 474, *a, b, c, d,* shows the different methods in which the field magnets are excited. These are (*a*) separately excited, that is the

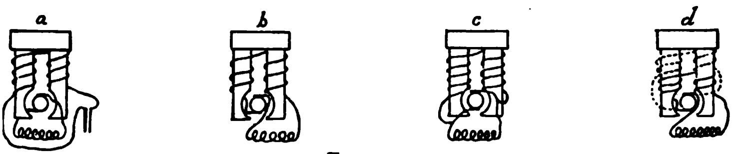


FIG. 474.

current in the field coils comes from a separate generator; (*b*) "**series wound,**" that is the field coils are in series with armature and the external circuit, so that all the current of the armature passes through the field coils as well as the external circuit; (*c*) "**shunt wound,**" that is, the field coils and the external circuit are in parallel, so that only a part of the current of the armature passes through the field coils; (*d*) "**compound wound,**" that is, part of the field coils are series and part shunt windings. The choice of windings of the field coils is largely a question of regulation of the electromotive force under different loads. For a discussion of these methods, the student must consult special manuals.

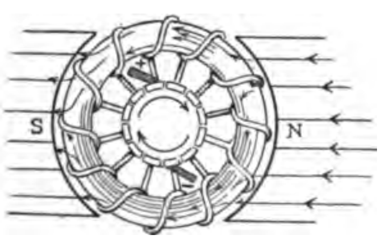


FIG. 475.

The two most common forms of D. C. armatures are: (*a*) the **ring armature,** sometimes called the Gramme armature, after its inventor, and (*b*) the **drum armature.** The ring armature is represented diagrammatically in Fig. 475. The coils are wound around a closed ring of soft iron, and connected as indicated. The core of the ring is laminated. Only the wires on the outside of the ring are inductors, as the wires on the inside are shielded magnetically. The course of the lines of force is indicated in the figure.

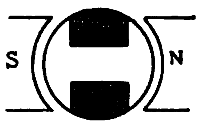


FIG. 476.

The first and the simplest form of drum armature is the "shuttle" armature used by Siemens in his machine of 1856. A section is shown in Fig. 476. The iron core increases the magnetic flux through the coil. This form is now used only in small magnetos. The winding and commutator connections of a moderr drum armature are very complicated. The coils may be on the surface or in tunnels or grooves in the core of the armature. The magnetic lines go from pole to pole as indicated in Fig. 477. For the study of the forms of armature and of their actions and reactions, the student must consult special treaties on dynamo-electric machinery.

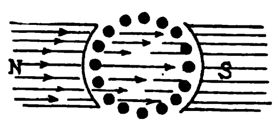


FIG. 477.

703. The Alternating Current Transformer. The alternating current transformer is a form of induction coil, used for transforming alternating currents of one potential into alternating currents of a different potential. It consists of a primary coil *Pr*, and a secondary coil *Sc*, and a laminated iron core to increase the magnetic flux. It is most commonly used to "step down" from a higher voltage to a lower voltage. The energy of the secondary current in well-designed transformers, is equal within a small percentage to the energy of the primary circuit. Thus a current of one ampere at 1000 volts is transformed into approximately 10 amperes at 100 volts. The *Pr* coil has in this case ten times the number of turns of the *Sc* coil. The only limit to the potentials that can be obtained with transformers is that of insulation. The coils of transformers for high potentials are generally immersed in a high insulating mineral oil.

704. Advantages of Alternating Currents in Power Transmission. Within recent years, electric power has been transmitted scores of miles, and alternating currents are used exclusively on these long distance power lines. The reasons for this general use of the alternating current in transmitting electric energy over longer distances are (*a*) the ease of transforming the alternating current from high to low potentials; (*b*) the possibility of securing high insulation in alternating current machinery; (*c*) the invention of the A. C. induction motor (§ 719).

Electric power is measured by the product of the current and of the potential, or equals Ie; thus the same power can be transmitted at a high potential with a small current, or at a low potential with a correspondingly larger current. But the weight of copper in the lines increases as the square of the current, since heating effects vary as I^2R (§ 667). There is thus a great economy in the transmission of electric power at high potentials, and indeed only in this way, is it commercially possible. Potentials of from 30,000 to 60,000 volts are in use for such transmission. But these high potentials cannot be used safely in lamps or in moving apparatus, so that it is necessary to transform to lower potentials before using the currents. For alternating currents, this can be easily and efficiently done by the A. C. transformer (§ 703). Such transformation is not possible with D. C. apparatus. The use of high potentials with alternating currents is also possible, because A. C. machinery can be insulated to stand the highest potentials. The A. C. dynamo has no commutator and the armature can in addition be made the stationary part. The A. C. induction motor (§ 719) also shares in the advantages of high insulation as well as efficiency and simplicity.

705. Two Phase and Three Phase A. C. Dynamos. The alternating current dynamo is often made so as to generate more than one alternating current. The alternating currents in such machines always differ in phase, and so they are known as polyphase currents. The only systems in commercial use

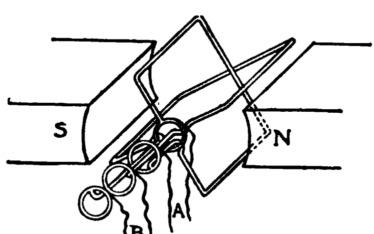


FIG. 478.

are the two and three phase systems. The currents are generated in coils placed at different angles on the armature. Thus in a two phase machine, there are two sets of armature coils, the first set cutting the magnetic field at a maximum rate, when the induced E. M. F. in the second set is zero, etc. Fig. 478 shows an arrangement of coils for a simple bi-polar machine by which two such currents could be generated. These currents differ in phase by a quarter period or 90°, and so a system using such currents is called a "quarter phase" system. Fig. 479 shows the phase of two

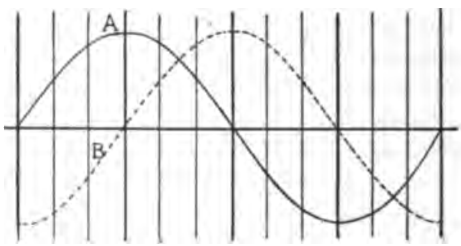


FIG. 479.

such currents. Fig. 480 represents the phase relations of the three currents, where the phase difference is 120°. A machine wound to generate only one alternating current is called a single-phase machine, to distinguish it from the polyphase machines. The frequencies used with two and three phase alternating currents,

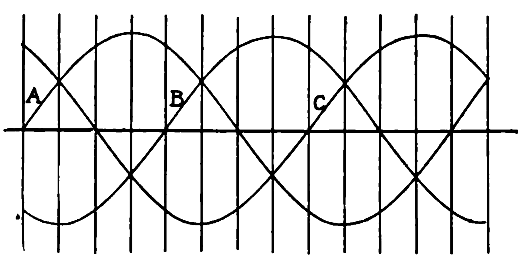


FIG. 480.

are the same as with single phase alternating currents, that is 60 alternations per second for ordinary conditions, with as low as 25 for power purposes alone. As already explained the com-

mercial machines are always multipolar for mechanical reasons.

The advantage of two and three phase alternating currents is that the induction motor can be used. A polyphase machine has also generally a larger output for weight of machine than a single phase machine. The relative advantage of the two and three systems, lies in the size of conductors required for the distribution of a given power. For a discussion of this, special treatises must be consulted.

706. Effect of Inductance in an A. C. Circuit. When an alternating E. M. F. acts in a non-inductive circuit, the current I and the E. M. F. are in the same phase, as is represented in Fig. 481. Here the current I at any instant is equal to e/R, where R is resistance of the circuit, and e is the E. M. F. in the circuit at the instant. When the E. M. F. is zero, the current is zero, and the maximum current corresponds in time to the maximum E. M. F., etc. If the circuit is inductive, then experiments with the oscillograph (§ 707) show that the current lags behind the external or impressed E. M. F., that is, the current reaches a maximum later than the impressed E. M. F. This lag is due to self-induction.

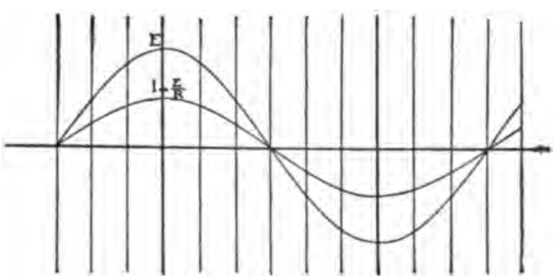


FIG. 481.

We have seen that starting or increasing a current in an inductive circuit produces an E. M. F. of self-induction which tends to retard the current growth, and similarly breaking or decreasing a current produces an E. M. F. of self-induction which tends to prolong the current. Thus the actual E. M. F. at any instant in an inductive circuit is the algebraic sum of the external or impressed E. M. F. (produced by generator, etc.), and the E. M. F. of self-induction. This E. M. F. of self-induction is equal to $-L(dI/dt)$ (§ 693).

In Fig. 482 we have the curves of the impressed E. M. F. E, of the effective E. M. F., which is equal to IR, and of the E. M. F. of self-induction $-LdI/dt$, represented in their phase relations.

In the coil of an electromagnet, where the inductance is large, the E. M. F. of self-induction is correspondingly large, and this may be sufficient to make the effective E. M. F. practically zero. Such a coil is called a *choking coil*, or an *impedance coil*. The resistance of an inductive circuit to an alternating current is called *impedance*. It can be shown that the impedance of a circuit is equal to $\sqrt{R^2 + 4\pi^2 n^2 L^2}$, where R is the resistance, L is the inductance of the circuit, and n is the frequency of the alternating current. The proof of this is given in treatises on alternating currents. This fact of impedance explains why little current goes through the primary of a transformer when the secondary circuit is not closed.

It is to be noticed that the impedance increases with the frequency n.

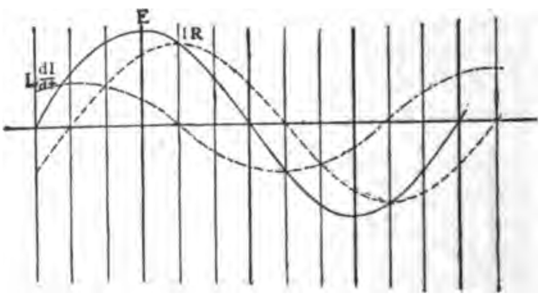


FIG. 482.

The frequency of a Leyden jar discharge is ordinarily very high (§ 722). and so a discharge has a large impedance even through a single loop. Thus in Fig. 483 the discharge will leap across a considerable air-gap P, sooner than go through the loop L.

707. The Oscillograph. An ordinary galvanometer shows no deflection from an alternating current, because the needle system has so much inertia that it cannot follow the rapid impulse from the alternating current. Blondel, Duddell, and others have made galvanometers with very light moving parts and of high frequency, so that the needle system can follow the changes in the alternating currents. A galvanometer with a high frequency needle system, so that its deflections show the variations in alternating currents,

is an oscillograph. Fig. 484 shows diagrammatically one of the best forms of oscillographs. It consists of a narrow loop of phosphor bronze strip, which is stretched with considerable tension, by a spring *s*, so as to have a very short natural period of

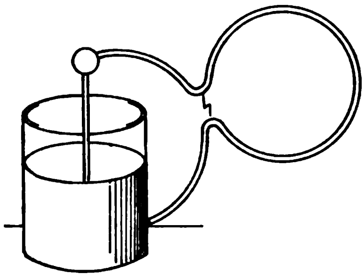


FIG. 483.

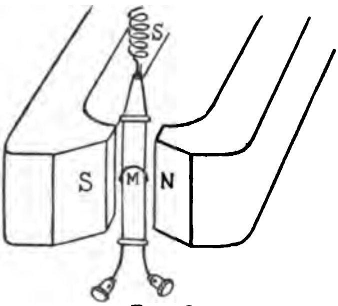


FIG. 484.

vibration. This is in the strong magnetic field and placed with the plane of the loop parallel to the magnetic field. The strip is thus twisted by the magnetic forces when a current passes through the loop. The natural frequency of the loop is commonly from 3,000 to 10,000 vibrations per second, so that its deflections can follow very closely all ordinary variations in alternating currents. The deflections are recorded by a beam of light which is reflected from a small mirror *M*, attached to the loop. This beam of light falls on a photographic plate which moves at right angles to the deflections. It thus leaves a curve showing the variations of the current.

Another form of oscillograph is the Braun cathode tube. The current passes through a helix, and thus acts magnetically (§ 732) on a pencil of cathode rays in a special form of Crookes tube (Fig. 485). When not deflected, this pencil of cathode rays produces a luminous spot on a phosphorescent screen in one end of the cathode tube. Under the action of an alternating current through the helix the luminous spot from the cathode rays vibrates in a luminous line on the phosphorescent screen. Looked at in a rotating mirror, this is drawn out into an alternating current curve. The same may be photographed on a plate moving at right angles to the vibrations of the luminous spot.

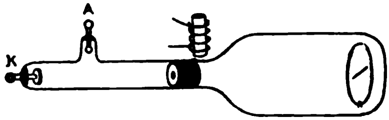


FIG. 485.

708. The Telephone. The telephone, invented by Bell in 1876, consists of a thin iron plate or membrane, supported in front of the pole of a permanent magnet, and a spool of wire over the magnet pole (Fig. 486). Sounds can be transmitted electrically to a distance by using two telephones, one for a transmitter, and the other for the receiver. The two wire spools are connected in series by the wires joining the two stations. The sound waves set the thin iron plate in vibration, and the approach or receding of this plate changes the magnetic flux through the coil. This induces currents in the coils and the line which undulate in unison with and in proportion to the sound waves. These currents strengthen and weaken the attraction of the magnet of the receiver, and thus produce vibrations of the receiver plate which correspond to the vibrations of the transmitter plate. The electric currents induced in the above cases are very feeble, and can transmit sounds only short distances. For longer distances, the microphone described in the next section is used as a transmitter.

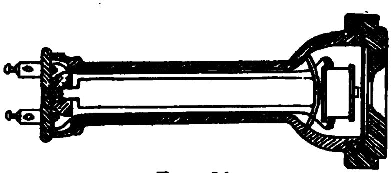


FIG. 486.

709. The Microphone. The microphone depends upon a fact discovered by Hughes in 1878, that the electrical resistance of a loose contact between two conductors changes under the action of sound waves. Variations of the current can thus be produced

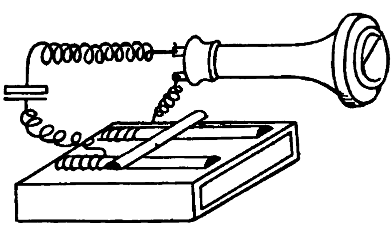


FIG. 487*a*.

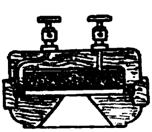


FIG. 487*b*.

in a circuit, these variations corresponding to the sound waves which produce them. A simple form of microphone consists of a piece of carbon resting on two pieces of carbon, and thus com-

pleting a circuit which includes a battery and a Bell receiver. The carbons can be mounted on a sounding box. Such an arrangement makes an effective transmitter (Fig. 487*a*). In the Hunning's transmitter, which has been extensively used in long distance telephony, granulated carbon is placed in between two metal plates as shown in Fig. 487*b*.

ELECTRODYNAMICS.

710. Motion of a Circular Circuit in a Magnetic Field. *Maxwell's Rule.* If a conductor carrying a current is placed in a magnetic field (§ 631), there is in general a force tending to move the conductor. It has been shown (§ 632) that a circular circuit or other plane closed circuit, acts like a magnet, and that the lines of force due to the current enter at the face in which the current flows clockwise, and emerge from the face in which the current flows anti-clockwise (Fig. 488). When such a coil is

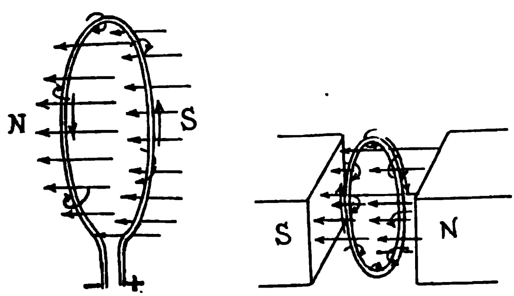


FIG. 488.

placed between the poles of a magnet, the coil tends to place itself, so that its plane is at right angles to the field, the clockwise face of the coil being toward the *N* pole; in other words, the coil places itself so that the lines of force of the field and of the coil coincide. Maxwell has generalized this into a rule— An electric circuit tends to move in a magnetic field so as to include the maximum number of lines of force. Thus the lines of the field and of the circuit are in the same direction when stable equilibrium is reached.

711. Force Between Two Parallel Circuits. If two circular coils carrying currents are hung by flexible wires parallel to each other (Fig. 489) they attract each other when the currents in the two coils are in the same direction, and repel each other, when the currents are in opposite directions. It is easy to see that each of the circuits thus tends to move so as to include the maximum number of lines of force. Rectangular or other plane closed circuits can be substituted for the circular circuits.

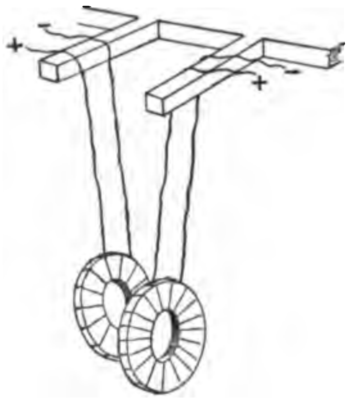


FIG. 489.

Lord Kelvin makes use of the forces between parallel circuits in his electrodynamic balance for measuring electric currents.

712. Force Between Two Circular Currents at Right Angles. Two circular currents at right angles to each other, tend to rotate and place themselves in the same plane, and with the currents in the same direction. Ampère, who was the first to study the actions of currents on currents, used his "electrodynamic apparatus" shown in Fig. 490. The two mercury cups *a* and *b*, are at the ends of the conducting supports *A* and *B*, and *a* is vertically over *b*. The wire frame *mnop* is bent so that its ends dip in the mercury cups, and the wire frame is thus free to rotate about a vertical axis through *ab*.

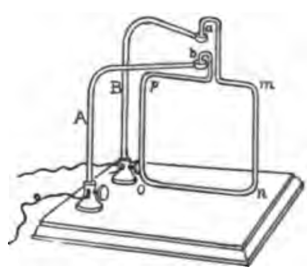


FIG. 490.

The force of rotation between coils is used in the electrodynamometer for measuring currents. Fig. 491 shows the Siemen's form of the electrodynamometer. The coils are kept at right angles to each other by putting more or less torsion in a spiral spring which is attached to the movable coil. The amount of torsion is read by a torsion head and circle on top. The force between the coils is proportional to the square

of the current, and the opposing force of torsion is proportional to θ, the angle of torsion. Hence we have $I^2 = k\theta$, or $I = k\sqrt{\theta}$, where k is a constant to be determined for each instrument. This instrument is adapted to the measuring of both direct and alternating currents.

FIG. 491.

713. Force between Parallel Linear Circuits. By using his electrodynamic apparatus, Ampère was able to show that two parallel straight circuits attract each other when the currents are in the same direction. An arrangement of the apparatus to demstrate this, is shown in Fig. 492. In Fig. 493 A and B are normal sections across the parallel conductors, and the circular lines of

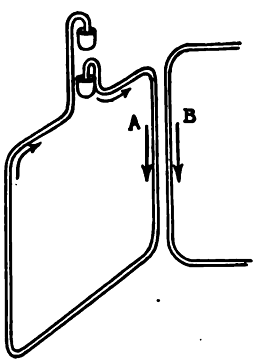


FIG. 492.

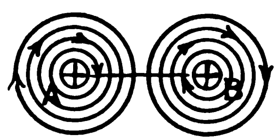


FIG. 493.

force are shown for the case of both currents flowing away from the reader. Evidently the maximum number of lines of force for each circuit exists when A and B are close together. Hence we should expect attraction from Maxwell's rule. The repulsion between oppositely directed currents follows similarly.

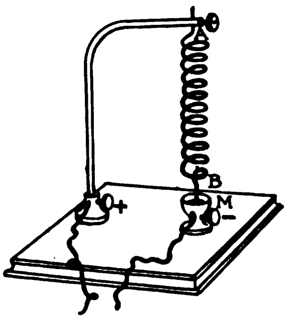


FIG. 494.

Ampère extended the above law as follows: Two straight conductors which cross each other obliquely, attract each other when their currents both flow towards or both away from the point of crossing; when one current flows away and the other towards the crossing point, they repel each other.

714. Roget's Spiral. An interesting case of the attraction between parallel circuits is that of Roget's spiral (Fig. 494). A spiral brass coil *AB* hangs vertically, and its end *B* dips in a mercury cup *M*. When a current passes through *AB* the parallel coils attract each other and lift *B*, thus breaking the circuit; the attraction ceases and contact at *B* is again made. This repeats itself indefinitely.

715. Electromagnetic Rotation. The first apparatus to produce rotation of a conductor from electric currents and magnets was described by Faraday in 1821. A glass tube (Fig. 495) is stopped

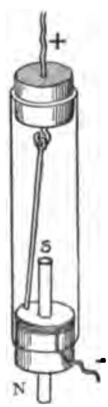


FIG. 495.

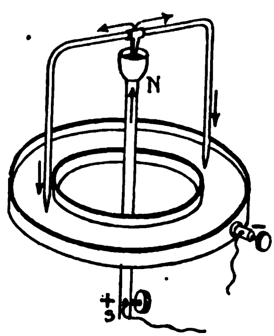


FIG. 496.

at both ends with corks, and the lower cork is covered with mercury, and has one pole of a magnet *NS* stuck through it. A platinum wire hangs by a hook from the upper cork and dips in the mercury below. When a current flows through the *platinum*

wire the wire revolves about the magnetic pole, and continues as long as the current flows. Reversing the current reverses the direction of the rotation. This is evidently a case of a conductor moving at right angles to the plane of the current and of the magnetic field from the pole (§ 631). Fig. 496 shows a common form of an electromagnetic rotation apparatus. Its action can be followed from the above.

716. Barlow's Wheel. This consists of a metal disk mounted to revolve about a horizontal axis over a mercury trough (Fig. 497). The lower edge of the disk touches the surface of the mercury. A U-shaped magnet lies so that its lines of force cut the lower half of the disk at right angles. When an electric current flows from the axis through the disk to the mercury, the disk rotates. Reversing the direction of either the current or of the magnetic field reverses the direction of rotation. A star-pointed wheel as shown in the figure is often used instead of the solid disk, in order to reduce the friction at the mercury surface. Each radius of the wheel in turn as it dips in the mercury becomes a part of the circuit, and is acted upon by a force at right angles to the plane of the magnetic field and the current (§ 631), and thus causes continuous rotation. Barlow's wheel and Faraday's disk (§ 690), are evidently inverse machines, the first using electrical energy to produce mechanical motion, and the second using mechanical energy to produce electrical energy.

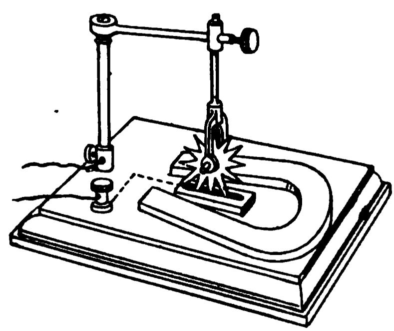


FIG. 497.

717. Direct Current Motors. The direct current dynamo becomes a motor when an external current is sent through its field magnets and armature. It then transforms electrical energy into the mechanical energy of the rotation of the armature. The forces acting on the armature circuits follow laws already treated in the sections on the motion of circuits in a magnetic field (§§ 631, 716).

The rotation of the armature in the magnetic field induces in

the armature coils an E. M. F. which opposes the current driving the machine. This back E. M. F. increases with the speed of the motor. Thus the current through the motor decreases as the speed increases. When the speed causes a back E. M. F. equal to the impressed E. M. F., there is no current through the armature. This occurs when a frictionless motor runs under no load, and is accordingly doing no work. At starting, there is no motion and no back E. M. F., and hence the current is a maximum. To prevent injury from the "rush" of the current before the motor reaches a speed to produce a back E. M. F., a starting resistance is commonly placed in series with the armature. This starting resistance is gradually reduced as the speed of the motor increases.

718. Alternating Current Motors. Two similar single-phase A. C. machines can be used as generator and motor, provided the motor is first brought to a synchronizing speed with the generator. An A. C. machine thus used as a motor is called a synchronous motor. The synchronous motor is efficient, but has the great disadvantage of not being self-starting, and also of stopping if the motor is thrown "out of step" by overloading. Hence synchronous motors are not in general use. The successful use of alternating currents for power transmission has been due to the invention of the polyphase induction motor. The principle of this motor was first discovered and stated by Ferraris, in 1885, and its application was developed by Tesla and others. Both two and three phase currents have been successfully used in these motors.

719. Simple Two Phase Induction Motor. In Fig. 498 we have represented two sets of helices AA' and BB', placed at right angles to each other.

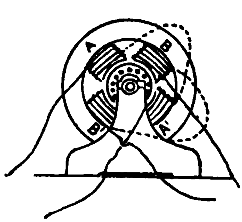


Fig. 498.

An alternating current through the helix AA' produces an alternating magnetic field in the line AA'. If the current through AA' is sinusoidal, represented by the equation $e_A = I \sin pt$ (§ 698), then the intensity of the field is represented by $h_A = NI \sin pt = H \sin pt$, where N is a constant depending upon the number of windings on AA', etc. Similarly, an A. C. in BB' will produce an alternating magnetic field in the line BB'. Suppose the A. C. in BB' to differ in phase from that in AA' by a quarter phase, that is, $i_B = I \sin (pt + 90°)$

$= I \cos pt$. Then the field BB' is represented by $h_B = NI \cos pt = H \cos pt$.

The field at any instant is thus the resultant of the two component fields, $_A = H \sin pt$, and $h_B = H \cos pt$. The resultant

$$F = \sqrt{h_A{}^2 + h_B{}^2} = \sqrt{H^2(\sin^2 pt + \cos^2 pt)} = H$$

That is the resultant is constant in strength. It evidently rotates with every complete alternation of the current.

Suppose we place in this rotating magnetic field, a "squirrel cage" armature, which can rotate about its axis OO', which perpendicular to the field. The construction of this "squirrel cage" armature can be seen from Fig. 499. It consists of two copper disks mounted on a shaft OO', and with copper bars around the circumference of the disks, joining the disks. It is found that such an armature rotates approximately synchronously with the rotating magnetic field. The explanation is simple. If the armature were held still, there would be electric currents induced in the bars. By Lenz's law these induced currents would react magnetically on the rotating field, with a force tending to prevent the relative motion of the armature and the field. In other

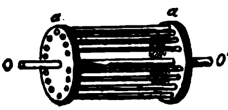


FIG. 499.

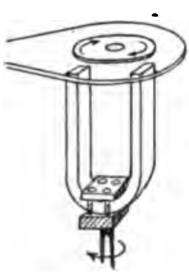


FIG. 500.

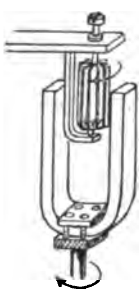


FIG. 501.

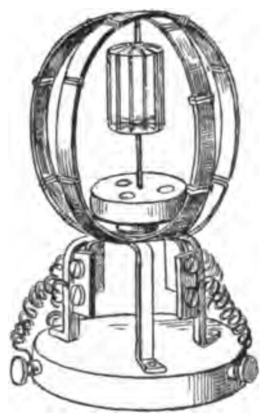


FIG. 502.

words, there is a torque causing the armature to rotate with the field. Figs. 500–502 show a series of simple experiments for demonstrating the principle of the rotating field, and the induction motor. In Fig. 500 we have an aluminum disk, mounted on a pivot. Under this is a magnet mounted on a whirling table so that it can be rotated. The disk ro-

tates with the magnet. This is the inverse of Arago's disk (§ 691), and the explanation by electromagnetic induction is the same. In Fig. 501, the experiment is varied by substituting a "squirrel cage" cylinder, pivoted on a vertical axis between the poles of the rotating magnet *NS*. In Fig. 502, we substitute for the rotating magnet, two coils, placed with the planes at right angles. By passing through the coils alternating currents, which differ in phase by a quarter period, a rotating magnetic field is produced, and the "squirrel cage" rotates.

ELECTRIC OSCILLATIONS AND WAVES.

720. Electric Oscillations. In the preceding sections, we have described alternating electric currents with frequencies which are commonly between 25 and 125 per second. It has been seen that these alternating currents follow special laws which are due to the special importance of inductance in such circuits. At still higher frequencies, alternating currents bring in new phenomena with additional laws. Alternating currents of high frequency are called oscillatory currents, or **electric oscillations.** The lowest frequency for which the term oscillatory is used is naturally not definite, but we may in general think of an electric oscillation as having at least 1,000 alternations per second. It is often several million per second.

The study of electric oscillations has been in recent years one of the most important and fruitful in physics. It has led to the discovery of electromagnetic disturbances in the space about oscillatory currents, disturbances which are propagated outward as electric waves. These electric waves have been shown to be identical physically with light waves, except in being of longer wave-length. Heinrich Hertz, the discoverer of electric waves, was thus able to prove experimentally the theory of James Clerk Maxwell, that light is an electromagnetic phenomenon (§ 724). The experiments of Marconi and others have resulted in using electric waves to transmit signals by the electric-wave telegraphy. Lodge and his fellow workers have also explained many of the "mysteries" of lightning discharges by laws proved for oscillatory currents.

721. Methods of Generating Electric Oscillations, Alternators. Two general methods of producing high frequency electric currents or oscillations have been used, (*a*) by multipolar alternating

dynamos, and (*b*) by an electric discharge in a circuit containing capacity and inductance in certain ratios, with low ohmic resistance.

A high frequency dynamo-electric machine must have a large number of poles, and be driven at a high velocity. Such machines have been constructed by Tesla, Ewing, Duddell and others. Frequencies of from 10,000 to 15,000 per second were ordinarily reached, and in one machine a frequency of 120,000 per second is recorded. But the velocity of the moving parts must be so high that only small machines are mechanically possible. The high frequency alternator is therefore not at present promising as a generator of electric oscillations.

The only methods of producing oscillations of the highest frequency are those based on the oscillatory character of the discharge of a circuit containing capacity. This will be described in the next section.

722. Oscillations by a Condenser Discharge. When a Leyden jar is discharged, there is a flash which to the eye appears as a single spark. But as early as 1842 Joseph Henry concluded that this discharge of a Leyden jar "is not correctly represented by the single transfer of an imponderable fluid from one side of the jar to the other." "The phenomena," he continues, "require us to admit the existence of a principal discharge in one direction, and then several reflex actions backwards and forwards, each more feeble than the preceding until equilibrium is obtained." Henry reached this striking conclusion by observing the irregular magnetization of steel needles by Leyden jar discharges. Henry's conclusion was confirmed by the mathematical theory of Lord Kelvin, published in 1853. Kelvin showed that the character of the discharge depended upon the resistance R, the capacity C, and the inductance L, and that the frequency is given by the equation,

$$n = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}.$$

If $R^2/4L^2$ is so small as to be neglected compared to $1/LC$, then the frequency is

$$n = \frac{1}{2\pi\sqrt{LC}},$$

or the period $T = 2\pi\sqrt{LC}$. That is, if the resistance of the discharging circuit is small, then the discharge is oscillatory. These oscillations are rapidly damped. When the resistance R is large, then the term under the radical is negative and the frequency be-

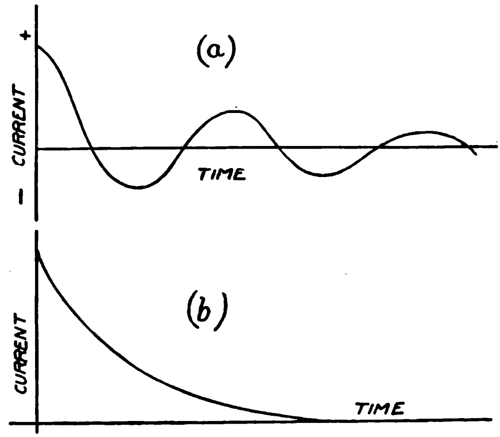


FIG. 503.

comes imaginary. The discharge is uni-directional, dying away slowly. Figs. 503, *a* and *b*, are curves showing these two types of discharge.

In 1858 Feddersen confirmed Kelvin's theory, showing by examining the spark discharge with a revolving mirror, that the spark consists of a series of alternating and diminishing flashes. Others have photographed these flashes. One of the most beautiful confirmations of the oscillatory character of Leyden jar discharge is shown in the photograph reproduced in Fig. 504. This was made by Zenneck in 1904, using a Braun tube as an oscillograph (§ 707).

FIG. 504.

The discharge of a condenser is thus analogous to the vibrations of an elastic rod clamped at one end. When bent and released, the rod in general vibrates back and forward about an

equilibrium position, dissipating its energy, and finally coming to rest. But if the rod is immersed in a heavy oil, which offers considerable resistance to motion, the rod comes slowly to rest without vibrating beyond its equilibrium position.

723. Electric Oscillations and Waves. Resonance. In 1888 Heinrich Hertz showed that a conducting system in which electric oscillations are produced becomes the source of electric waves, and that these waves can be detected by oscillations set up in a similar circuit called a resonator. Fig. 505 shows one of Hertz's arrangements. The discharge rods *A* and *B* are connected to terminals of the secondary of the induction coil *C*, and are separated by the discharge gap *P*. The metal spheres *S* and *S'* slide on the rods, so that the length of the discharge circuit can be varied. For a receiving circuit or resonator, Hertz used a loop of wire *R* broken by the spark gap *P'*. He found that when the two circuits were

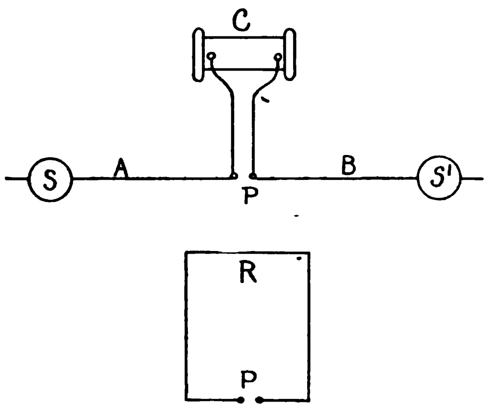


FIG. 505.

"in tune," a discharge at *P* caused a spark at *P'*; or in other words, oscillations in the first circuit produced oscillations in the second circuit. The explanation is evidently exactly like that of the experiment of resonance between two tuning forks on resonators. The sound waves sent out from a tuning fork *A* set in vibration a second fork *B*, provided the two forks are of the same pitch. The electric waves from the oscillator produce the electric oscillations in the resonator, provided they are "in tune." Fig.

506 shows a striking class-room experiment due to Lodge for showing electrical resonance. *A* and *B* are two equal Leyden jars. The jar *A* has a wire loop *L* which forms the discharge circuit, the gap being between the polished balls at *P*. The jar is charged by a small static electric machine. The inner and outer coatings of the jar *B* are connected by a wire loop *L'*, the inductance of which can be varied by the sliding wire *m*. By using a tin-foil strip, a small gap *G* is left between the inner and outer coatings of *B*. When the two circuits are in tune, a discharge in *A* produces oscillations in *B*, which are shown by a bright spark at *G*.

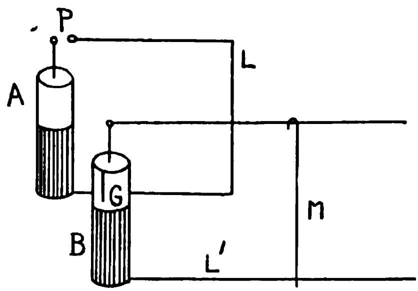


FIG. 506.

724. Electromagnetic Theory of Light. Using his spark gap detector, Hertz showed that electric waves are reflected from plane and curved metal surfaces in accordance with the same laws as light waves; that they are refracted in passing through prisms of resin, paraffin and other dielectrics; that they are polarized by a coarse metal grating, and hence are transverse waves. He measured their wave-length and computed from his oscillator their frequency; and thus, from the formula $v = n\lambda$, he determined that the velocity of electric waves is the same as that of light. The electric waves, which Hertz produced, generally had wave-lengths of eight or nine meters. The shortest electric wave yet produced has a wave-length of about four millimeters, still many times the length of the longest infra-red line (§ 510).

Twenty years before Hertz's experiments were performed, Maxwell advanced the view that waves of light are electromagnetic waves of very short wave-length. From theoretical calculations Maxwell found that the velocity of such waves equals $1/\sqrt{k\mu}$, where k is the dielectric constant of the medium and μ its permeability, both being expressed in electromagnetic units. The velocity thus calculated for air agrees with the velocity of light (§ 729). The value of μ for transparent substances is nearly 1.

Hence the index of refraction (§ 442) from a substance of dielectric constant k_1 to another of dielectric constant k_2 is $n = \sqrt{k_2/k_1}$. This relation has also been verified in many cases, but the dependence of n on the wave-length makes the test difficult in other cases. The waves started from Hertz's oscillator (§ 723) are *plane polarized.* At P' there is an alternating electrostatic force in the plane of the diagram and an alternating magnetic force perpendicular to that plane. These together constitute the vibration in the front of the wave, and a plane polarized wave of light is similarly constituted. Thus the electromagnetic theory completes the wave theory stated in Light, by explaining the nature of the wave-motion.

725. Electric Waves along Wires. Fig. 507 shows a form of

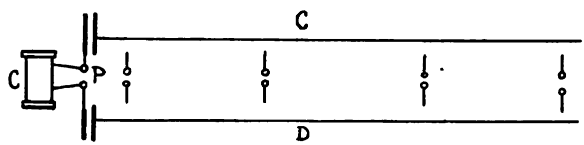


FIG. 507.

Hertz oscillator as modified by Lecher to show electric waves along wires. The oscillations produced by the discharge across P, act by static induction, and produce waves which traverse the wires C and D and are reflected back, thus forming standing waves by interference between the advancing and the reflected waves (§ 355), similar to the standing waves in organ pipes (§ 382). The nodes and loops can be detected by sliding a small gap along the wires, or easier by a device due to Arons, shown in Fig. 508.

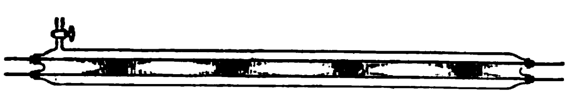


FIG. 508.

Arons enclosed the two wires in an exhausted glass tube. The loops are indicated by the electrical discharges, while the nodes remain dark.

Seibt has arranged a beautiful class-room experiment (Fig. 509) in which he uses a Tesla transformer T (§ 696) as oscil-

lator, and a special resonance coil *CD* to show standing waves. The vertical coil *CD* is about two meters high and consists of a coil of silk-covered wire on a wooden core. Parallel to it and insulated from it, is a stretched wire *MN*. The nodes and loops come out brilliantly in a darkened room, as indicated in Fig. 510.

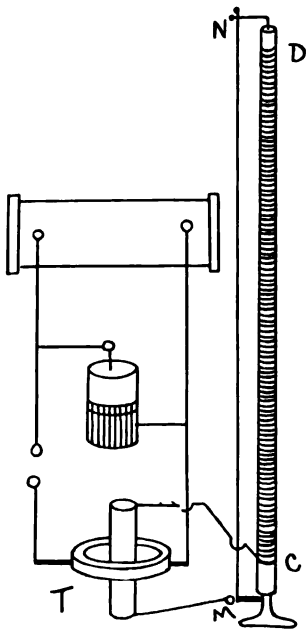


FIG. 509.

FIG. 510.

726. Detectors of Electric Waves. The spark gap, which Hertz used so successfully in his investigations, has been largely replaced by more sensitive detectors. Cymoscope has been proposed as a general name for electric wave detectors. Almost every effect of an electric current has been used in these detectors, such as heating, magnetic, electrolytic, and resistance effects. Only one of these detectors, the coherer, will be described here. The reader is referred to special treatises for accounts of the others.

The coherer, in the form given to it by Marconi, consists of a small glass tube *TT′* (Fig. 511), in which there are two silver electrodes *PP′*, separated by a small quantity of loosely packed metal

filings. A mixture of 95 per cent. nickel and 5 per cent. silver filings has been successfully used by Marconi. Marconi also found that exhausting the tube of air increased the reliability of the coherer. The action of the coherer depends upon a dis-

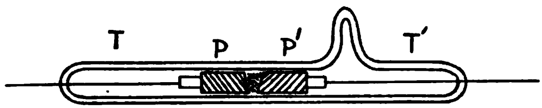


FIG. 511.

covery made by Branly in 1900. He discovered that loosely packed metal filings, which offered practically infinite resistance to an electric current, suddenly acquire good conductivity under the action of an electric wave. When lightly tapped or shaken, the filings again lose their conductivity. The generally accepted explanation is that the small filings cohere owing to the welding action of the infinitesimal sparks produced by the electric wave, and hence the name coherer was given. The coherer is not selective in its action, that is, it responds to electric waves of many or all lengths. The method of using the coherer can be seen from the diagram in the next section.

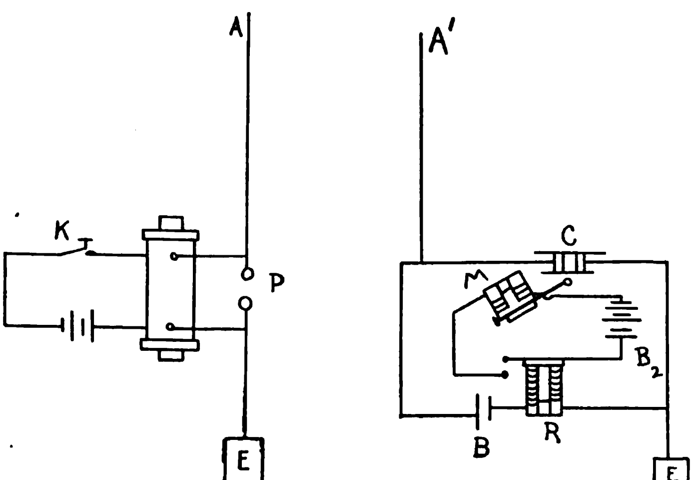


FIG. 512.

727. Electric Wave Telegraphy. Since 1895 Marconi has developed a system of electric wave telegraphy, more often called wireless telegraphy, for transmitting signals to a distance. Using very powerful oscillators and extremely sensitive detectors, Marconi has transmitted messages thousands of miles. This system has been particularly successful in communicating with and between ships at sea. Fig. 512 shows a diagram of an electric wave telegraphic arrangement. A and A' are high vertical lines. P is the spark gap of the sending station, C is the coherer, R is a relay operated by any current through C. This throws in the battery B_{ii} and excites the magnet M, which decoheres C by tapping it. E and E are earth connections.

DIMENSIONS OF ELECTRICAL UNITS.

728. Kinds of Electrical Units. Three kinds of electrical units have been defined and used in the previous sections, the electrostatic units, the electromagnetic units, and the "practical units." The practical units have been defined as multiples of the electromagnetic units, the multiples being chosen so as to make units of convenient sizes for calculations in the technical applications of electricity. The electrostatic and electromagnetic units are both "absolute units," that is are based by definitions on simple relations to the fundamental units, the units of length, mass, and time (§ 150). The particular absolute system long universally used in electricity and magnetism is that based on the centimeter, the gram, and the second, or the C. G. S. system (§ 150).

The establishment and universal use of an absolute system of units in electricity and magnetism has contributed much to the progress of the science both in its theory and in its applications. The relations of the units of electric quantity, current, potential, etc., to the units of energy and power are clear and direct in an absolute system. Thus the product of the number of units of current and of potential gives directly the number of units of power or activity, no arbitrary constants entering into the calculations. The advantage of this simplicity is evident. Again the study of the dimensions of the units (§ 151), has led to a clearer view of the nature of electrical and magnetic quantities, and of the relations of electrical phenomena to other phenomena. Thus the

comparison of the dimensions of the electrostatic and electromagnetic units suggested to Maxwell important similarities of the electrical and optical ethers, and contributed much to Maxwell's electromagnetic theory of light (§ 724). This last theory was again a starting point for speculations which resulted in Hertz's epoch-making experiments on electric waves and their properties (§ 723). The dimensions of electrical and magnetic units thus have a greater importance than that of translating results from one absolute system to another (§ 151).

729. Dimensions of Electrical Units. The following table gives the dimensions of five of the more usual electrostatic and electromagnetic units.

Name.	Symbol.	Electrostatic.	Electromagnetic.
Electric Quantity	q	$[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Magnetic Quantity	m	$[L^{\frac{1}{2}}M^{\frac{1}{2}}k^{-\frac{1}{2}}]$	$[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$
Magnetic Field	H	$[L^{\frac{1}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{-\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Current	I	$[L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$	$[L^{\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$
Potential or Electromotive Force	$\begin{cases} V \\ E \end{cases}$	$[L^{\frac{1}{2}}T^{-1}M^{\frac{1}{2}}k^{-\frac{1}{2}}]$	$[L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$

The method of deriving the above dimensions from the definitions is shown by the following examples.

Electrostatic Unit of Quality. We have by definition (§ 583) $q = r\sqrt{Fk}$. Using the dimensions of r and F (§ 154), we get $[q] = [L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}k^{\frac{1}{2}}]$. In this k is the specific inductive capacity or dielectric constant (§ 583), a quantity arbitrarily assumed as unity for air but of undetermined dimensions.

Electrostatic Unit of Current. By definition of (§ 621) $[I] = q/t$. Substituting the dimensions, we get $[I] = [L^{\frac{3}{2}}T^{-2}M^{\frac{1}{2}}k^{\frac{1}{2}}]$.

The starting point in the electromagnetic system is the definition of unit magnetic pole (§ 557), $m = r\sqrt{F\mu}$, where μ is the magnetic permeability (§ 565), a quantity arbitrarily assumed as unity for air, but of undetermined dimensions. From this we get the dimensions of $[m] = [L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$. From the relation that F, the force at a point in a magnetic field is mH, we get $H = F/m$. The dimensional equation for intensity of magnetic field is thus $[H] = [L^{-\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{\frac{1}{2}}]$.

Electromagnetic Unit of Current. The strength of magnetic field at the center of a circular coil of radius r, and carrying a current I, is $H = 2\pi I/r$ (§ 629); substituting dimensions, we get $[I] = [H][L] = [L^{-\frac{1}{2}}T^{-1}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$.

Electromagnetic Unit of Quantity. From the relation $q = It$, we get $[q] = [L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$.

Comparing the electrostatic and electromagnetic units of quantity, we get the ratio $[L^{\frac{3}{2}}T^{-1}M^{\frac{1}{2}}\kappa^{\frac{1}{2}}] \div [L^{\frac{1}{2}}M^{\frac{1}{2}}\mu^{-\frac{1}{2}}] = [LT^{-1}\kappa^{\frac{1}{2}}\mu^{-\frac{1}{2}}]$. But LT^{-1} is a velocity (§ 154). This velocity "v" also appears in the ratios of the other units, though not always as the first power. If an electric quantity is measured in air both electrostatically and electromagnetically, then both k and μ are assumed as unity and the value of this velocity "v" can be determined. This was first done by Weber and Kohlrausch in 1856, by determining the electric quantity in a condenser from its electrostatic capacity and potential (§ 591), and also by discharging the same quantity through a ballistic galvanometer (§ 639). They obtained the value $v = 3{,}107{,}040{,}000$ meters per second. This number is within limits of error the same as the velocity of light. This equality has been established by numbers of later determinations. The close connection between the velocity of light and the ratio of the electrostatic and electromagnetic units, confirmed Maxwell in the theory that light is a phenomenon of the same ether as that of electromagnetic actions (§ 724).

If we assume the equality of the two units of quantity, without assuming k and μ as unity, we get directly that "v" $= 1/\sqrt{k\mu}$. This is a very significant relation, and has been the subject of much experiment but has been only partially confirmed.

REFERENCES.

Faraday, Michael. *Experimetal Researches in Electricity.* (Historical Interest and Experimental Method.)

Gray, Andrew. *Theory and Practice of Absolute Measurements.* (Theory of Measurements.)

Thomson, J. J. *Mathematical Theory of Electricity and Magnetism.* (Electro-statics.)

Whetham, W. C. D. *Theory of Experimental Electricity.* (Good general text.)

Franklin, Williamson. *Elements of Alternating Currents.* (A. C. Machineny.)

Fleming, J. A. *Principles of Electric Wave Telegraphy.* (Wireless Telegraphy, etc.)

Watson. *A Textbook of Practical Physics.* (Laboratory manipulations.)

Maxwell. *Electricity and Magnetism.* Vol. 2.

Poincaré. *Maxwell's Theory of Wireless Telegraphy.*

Stewart. *Electricity.*

PROBLEMS.

1. Find the intensity of field at a point 40 cm. from a magnet in the perpendicular bisector of the line joining the poles of the magnet 6 cm. long and of pole strength 160 E. M. U. Calculate the force on a pole of + 80 E. M. U. if placed at the point.

Magnetic Field.

Ans. .0148 E. M. U.; 1.19 dynes.

2. Derive formula for field strength at a point in the line through the poles of a magnet.

$$H_a = \frac{2mld}{\left(d^2 + \frac{l^2}{4}\right)^2}.$$

Electrostatic Field.

3. Two small spheres, each weighing 1 decigram, having equal charges, are suspended from the same point by silk fibers 80 cm. long. If the spheres are kept 8 cm. apart by repulsion what is the charge on each? Ans. 17.69 E. S. U.

4. Given two spheres of radii, 3 cm. and 8 cm., how will a charge of 66 units distribute itself over them if they be connected by a fine wire? Ans. 18 and 48.

5. Two charges + 90 and — 40 are 30 cm. apart. Find the intensity of field at a point in the line joining them 60 cm. from the negative and 90 cm. from the positive charge, and calculate the force on a charge of + 20 if placed at this point. Ans. 0; 0.

6. A charged pith ball of negligible size is fastened to a perfectly smooth insulating plane of slope $\sin^{-1} 0.15$. Directly up the plane 5 cm. a pith ball of mass 0.5 gram charged with + 21 E. S. U. is just kept from rolling down. What is the charge on the fixed ball? Ans. 87.5 E. S. U.

Capacity.

7. A Leyden jar ¼ cm. thick is 3 cm. in radius and 9 cm. high. Find its capacity, if the dielectric constant for glass is 6. Find charge on each plate when p.d. is 15 E. S. U. Ans. 378 E. S. U.; 5670 E. S. U.

8. A condenser of 10 plates, each 20 cm. × 30 cm. has 0.4 mm. of air between each pair of plates. Find the capacity. Ans. 10740 E. S. U.

9. Two plate condensers are joined in parallel. One is a 15 plate air condenser, each plate 11 cm. long and 5 cm. broad, 3 mm. apart; the other a mica condenser of 10 plates, 22 cm. long, 15 cm. broad, 0.5 mm. apart, specific inductive capacity of mica being 8. Find the capacity. Ans. 38035 E. S. U.

10. Two concentric spheres of radii 10 cm. and 10⅓ cm. are separated by air and are charged to difference of potential of 50 volts. Find charge. Ans. 51.67 E. S. U.

11. A pair of circular plates of radii 10 cm. each are 2 mm. apart in air. They are charged to a difference of potential of 20 and are then connected to the plates of an uncharged condenser and the difference of potential falls to 3. Find the capacity of this condenser. Ans. 708.33 E. S. U.

12. Find the capacity of a plate condenser made of two rectangular conductors 32 cm. long and 22 cm. broad, 0.2 cm. apart in air. Ans. 280.11 E. S. U.

13. If the air be replaced by 0.2 cm. sheet of glass of dielectric constant 7, find the charge on each plate when the difference of potential is 20 E. S. U. Ans. 392.16 E. S. U.

Specific Resistance.

14. Calculate the resistance of 1.10 kilometers of copper wire of diameter 0.84 cm., taking specific resistance of copper to be 0.0000018 ohms cm.^{2} per cm. Ans. .35729 ohms.

Ampere's Formula.

15. Calculate current which will deflect a tangent galvanometer 45°, if the galvanometer consists of a coil 18 cm. in diameter, of 7 turns of wire, set up in a field of 0.198 lines per cm^2. Ans. .4052 amp.

16. Find the field strength 16 cm. from the center of a coil in the line of its axis if the coil carry 0.5 amp. and be 24 cm. in diameter. Ans. .00565 E. M. U.

17. Find force on a pole of 30 E. M. U. if placed at center of coil.

Ans. .7854 dynes.

Ohm's Law and Joule's Law.

18. Given 3 cells of 1.4 volts and 0.8 ohms resistance each, find resistance of the battery if the cells be connected in series and calculate the current through an external resistance of 9 ohms.

Ans. (*a*) 2.4 ohms; (*b*) .36842 amp.

19. Find the resistance of the battery if cells be in parallel and also current through 9 ohms external resistance. Ans. .2667 ohms; .1511 amp.

20. By experimenting with a Weston ammeter it was found that 0.00013 amperes through the coil gave one unit scale-deflection. If the resistance of the coil circuit be 5.60 ohms what must be that of the shunt so 1 ampere in the external circuit will give 1 unit scale deflection?

Ans. .000728 ohms.

21. It is desired to supply 600 incandescent lamps, in parallel, with ½ amp. each, at 110 volts potential difference between the lamp terminals. If the drop in the line be 2.2 volts what is the resistance of the line and how much power is lost in it? How much power must be generated and what voltage? Ans. .007333 ohms; 660 watts; 3.366 K. W.; 112.2 volts.

22. A car is lighted by five lamps of 220 ohms resistance each, joined in series. What is the total resistance of the lamps? If the difference of potential between the ends of the lamp circuit be 550 volts, what current flows through the lamps? What power is expended in this circuit and at 9c. per K. W. hr., what does it cost to light a car for one hour?

Ans. 1100 ohms; .5 amp.; 275 watts; 2.475 c.

23. If the motive circuit of a snow sweeper take 50 amp. (at 550 volts) and the broom motors take 80 amp., find the total power consumed in the car if the two circuits be in parallel across 550 volt mains. Find cost per hr. at 9c. per K. W. hr. Ans. 6.435 c.

Electromagnetic Induction.

24. Show that Lenz's law and Fleming's rule lead to the same direction for an induced current in a conductor moved across a magnetic field.

25. The diameter of a circular coil is 30 cm. and the resistance is 0.1 ohm. Find the quantity of electricity in coulombs which will flow in the ring when revolved from a position at right angles to a magnetic field to a position parallel to the field. $H = 20$. Ans. .001414 coulombs.

26. A circular coil 40 cm. in diameter and with 100 turns is rotated five times per second about a vertical axis. Find the maximum E. M. F. induced. The horizontal component of the field is 0.2. Ans. .007893 volts.

27. If the angle of dip is 70°, what is the maximum E. M. F. induced when the above coil is rotated about a horizontal axis parallel to the horizontal component of the field ten times per second? Ans. .02169 volts.

28. Calculate the E. M. F. induced in a car axle length 120 cm. and with a velocity of 25 meters per second, where the total intensity of the field is 0.6 and the angle of dip is 70°. Ans. .00169 volts.

29. Calculate the number of revolutions per second which must be given to a disk of 60 cm. diameter to produce an E. M. F. of 5 volts between the center and the periphery of the disk, the axis of the disk being parallel to the field, and the field being uniform and of strength 10,000.

Ans. 17.68 r. p. s.

30. Draw a figure showing the directions of the induced currents in the disk of a pendulum swinging between the poles of a magnet across the field. How should the disk be laminated to make the induced currents a minimum?

CONDUCTION OF ELECTRICITY THROUGH GASES AND RADIO-ACTIVITY.

By R. K. McClung, D.Sc.,

Professor of Physics, Mount Allison University, Sackville, N. B.
Late Senior Demonstrator in Physics, McGill University.

CONDUCTION OF ELECTRICITY THROUGH GASES.

730. Introduction. Air, as well as other gases, under normal conditions is almost a perfect non-conductor of electricity. When a difference of potential is established between two points in a gas the gas is in a state of strain, as has been explained in a former paragraph (§ 587). This strain increases with increase of potential until, when a certain potential is reached, the air is no longer able to withstand the strain and breaks down and a discharge passes. A momentary current of electricity is thus produced through the gas. To produce such a discharge a comparatively large potential is required, several thousand volts being necessary to produce a spark of 1 cm. length in air at atmospheric pressure. The potential necessary to produce a discharge depends upon the shape of the electrodes and the nature and pressure of the gas.

731. Effect of Pressure of a Gas on the Discharge. If two metal electrodes are inserted in the ends of an air-tight glass tube, such as shown in Fig. 513, filled with air at atmospheric pressure, and if sufficient voltage is applied to the electrodes the discharge ordinarily obtained in air will be observed. If the air be gradually exhausted from the tube the discharge will pass with greater and greater ease as the pressure is diminished, until a certain minimum pressure is reached, and if the exhaustion be carried beyond this point the voltage necessary to produce a discharge will increase somewhat rapidly, until at the lowest pressure obtainable it will be impossible to cause

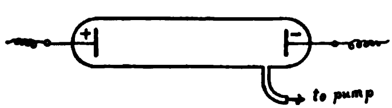


Fig. 513.

a discharge to pass at all. The pressure corresponding to this minimum potential is called the critical pressure and varies with the distance between the electrodes.

As the pressure is gradually diminished below atmospheric pressure the appearance of the discharge changes very much. At first the spark becomes more regular and uniform between the electrodes, then broadens out and assumes a fuzzy appearance of a bluish color. When a pressure of about half a millimeter is reached the discharge assumes a very marked appearance, which is shown in Fig. 514. The surface of the negative electrode or

FIG. 514.

cathode is covered by a very thin layer of luminosity; next to this is a dark space which is called the Crookes dark space; immediately beyond this dark space is a luminous part called the negative glow, and then beyond this again is a second dark region, sometimes called the Faraday dark space. Between this and the anode there is a luminous region which goes under the name of the positive column. Under certain conditions of current and pressure the positive column shows alternately dark and light spaces which are called *striæ*. The position of the positive electrode, or anode, in the tube does not affect the position of the negative glow, for if the anode be placed in a side tube as in Fig. 515, the positive column will bend into this tube to meet the anode, but the negative glow may extend in a straight line beyond this side tube. The proportion of the space between the electrodes occupied by each of these sections of the discharge depends upon the distance between the electrodes. Any increase in the length of the discharge beyond a few centimeters causes an increase in the length of the positive column but no increase in the negative glow or dark space. Similar phenomena occur in other gases besides air.

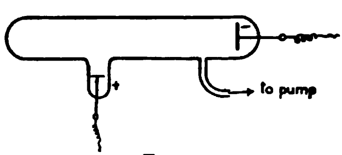


FIG. 515.

732. Cathode Rays. When the pressure in such a discharge tube is lowered to the neighborhood of a hundredth of a millimeter, a new phenomenon makes its appearance. The positive column begins to disappear and a bright phosphorescence appears on the sides of the tube. This phosphorescence appears to be produced by radiations or streams of very minute particles issuing normally in straight lines from the cathode. They are, consequently, called *cathode rays,* and possess remarkable properties.

They were discovered in the year 1859 by Plücker, who observed the phosphorescence produced by them. He also observed that if he brought a magnet close to the tube the rays were deflected from their original path. A few years later Hittorf discovered that if he placed a solid body inside the tube in the path of the rays it casts a well defined shadow. Numerous other substances besides glass are caused to phosphoresce by the impact of cathode rays. For instance, such substances as calcspar, potassium platino-cyanide and some of the rare earths show this effect.

The bombardment of some substances by cathode rays causes a marked heating effect. If the rays be concentrated upon a platinum plate inside the tube the plate may be heated to incandescence.

Chemical changes are also produced in some substances by cathode rays, as is shown by the change of color of the substance acted on. This effect is usually not a permanent one.

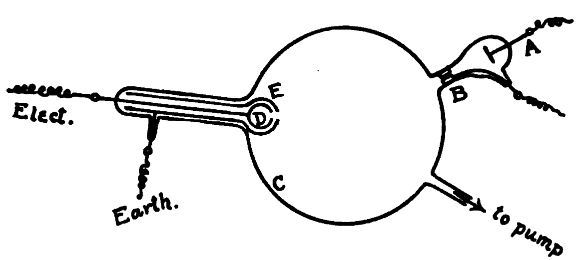


FIG. 516.

One of the most important properties of the cathode rays is that they carry a negative charge of electricity. This was originally proved by Perrin and his method was later modified by J. J. Thomson. A diagram of the apparatus used in the latter

experiment is shown in Fig. 516. A was the cathode and B the anode. The cathode rays from A passed into the larger part of the tube through a hole in B and fell upon the glass at a point C. A side tube contained two coaxial metal tubes. The outer one E had a slit in the end and was connected to earth. This shielded the inner tube from any stray electric effects. The inner tube D had a slit opposite that in E and was insulated from E and connected to an electrometer. When the cathode rays were allowed to fall upon the glass bulb the electrometer indicated only a very small effect, but if the rays were deflected by means of a magnet so that they fell upon the slits in the cylinders D and E the electrometer indicated that the cylinder D had received a considerable negative charge. If the rays were deflected still further so as to miss the slit the cylinder immediately ceased to receive any charge. This experiment clearly shows that the rays are accompanied by a negative charge of electricity. If the cathode rays be allowed to pass between two parallel plates inside a highly exhausted cathode ray tube, such as is indicated by Fig. 517, and a large difference of potential be

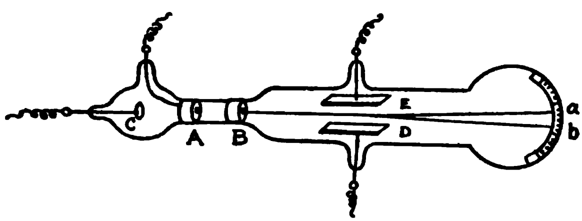


FIG. 517. (After J. J. Thomson, Conduction of Electricity through Gases.)

established between the plates, the beam of rays will be deflected and the deflection will be in the same direction as a negatively charged particle would be moved by the field.

The facts that these rays may be deflected by a magnetic and also by an electric field and that they are accompanied by a negative charge point strongly to the conclusion that the rays consist of small negatively charged particles shot out in straight lines with a high velocity from the cathode. This theory is now universally accepted to explain the nature of these rays and is well supported by experiment.

733. Velocity, and Ratio of the Charge to the Mass, of a Cathode

Ray Particle. We will now consider the method which J. J. Thomson originally used to determine experimentally the velocity of these particles and the relation between the mass of a particle and the charge which is carries. A highly exhausted cathode ray tube was arranged as shown in Fig. 517. *C* was the cathode, *A* the anode, and *B* a thick metal plug. *A* and *B* were pierced by holes in the same straight line about a millimeter in diameter, so that a very narrow beam of rays might pass along the middle of the tube and fall upon a screen of phosphorescent material, thereby producing a small bright spot. *D* and *E* were two parallel plates which could be connected to the poles of a battery. Suppose that V is the velocity of the particle in cms. per sec., m its mass, and e the charge which it carries, measured in electromagnetic units. If the tube be placed between the poles of a strong electro-magnet, so that a field of strength H is acting at right angles to the beam, the spot on the screen will move from a to b in a direction at right angles to the lines of force. The cathode particle will follow a curved path just as a moving projectile follows a curved path when acted on by gravity. Let the radius of curvature of this path be r. The deflecting force acting along this radius of curvature is proportional to the magnetic field, the charge on the particle, and its velocity and is, consequently, equal to HeV (see § 622). This force must equal the centrifugal force of the particle which, from dynamics, is equal to mV^2/r (see §§ 33, 42). Therefore,

$$HeV = \frac{mV^2}{r}$$

$$\therefore Hr = \frac{mV}{e} \qquad (1)$$

H can be measured and r may be found from ab and the dimensions of the apparatus. Therefore the quantity mV/e is known. Suppose now that a difference of potential be established between the plates; an electric force will act on the beam of rays and if it is applied in the right direction it will tend to deflect the beam in a direction opposite to the magnetic deflection. Suppose the magnetic and electric forces be adjusted so that their effects on the particles exactly balance each other, then the phosphorescent spot will return to the position it had before any force acted on it. Let this electric field be X. The force acting on the particle

will then be Xe and, therefore, if the electric and magnetic forces exactly balance each other

$$Xe = HeV$$
$$\therefore \quad V = X/H \tag{2}$$

X and H can both be measured and, therefore, V may be determined, and knowing V the value of e/m is easily found from equation (1).

By this method Thomson found the value of V to be 2.8×10^9 cms. per second, which is just about one tenth the velocity of light. This value is not quite constant as it varies somewhat with the potential in the tube. He also found a value for e/m the magnitude of which, according to later determinations, is 1.7×10^7; and he discovered that it was independent of the nature of the gas in the tube.

The greatest value of e/m known in electrolysis is found in the case of the hydrogen ion and is about 10^4. The value for the cathode ray particle is thus 1700 times that for the hydrogen ion. In a later paragraph (§ 745) the charge e carried by the cathode particle will be determined and it may be shown to be the same as for the hydrogen ion. Consequently, the mass of the cathode particle must be about 1/1700 of the mass of the hydrogen ion or atom. This cathode particle possesses the smallest mass yet known, and to it Professor Thomson has given the name of negative "corpuscle." By others it is sometimes called an "electron," while the term negative "ion" is also applied to it.

734. Lenard Rays. It was long considered impossible for cathode rays to penetrate any solid material. Hertz was the first to show the error of this, and later Lenard made a thorough investigation of the question. He used a special cathode ray tube, the end opposite the cathode being made of metal, and through this a small hole was bored and covered with very thin aluminum foil. When the rays fell upon this window a distinct phosphorescence was produced just outside the window. If the air outside this aluminum window was at ordinary pressure the rays did not proceed far, being soon absorbed by the air, but if the window was covered with another tube from which the air could be exhausted, the rays formed a distinct beam in the gas at low pressures and acted in all respects exactly like cathode rays. Lenard determined the value of e/m for these rays and found it to be the same as for cathode rays. These rays are identical with the cathode rays, but to distinguish them from those produced inside the cathode ray tube they are called Lenard rays.

735. Canal Rays. Goldstein in working with a highly exhausted tube observed that if he used a perforated cathode instead of a solid one luminous streams emerged through the hole in the cathode in the direction opposite to the cathode rays. These rays, which have been called canal rays, produce phosphorescence and may be deflected by a magnetic or electric field. The direction in which they are deflected shows that they carry a positive charge. W. Wien has determined the value of e/m for these positively charged particles, and the greatest value which he obtained was 10^4, which is the same as the value of e/m for the hydrogen atom in electrolysis. This indicates that the mass of these positive ions, as they are called is of the same order as the mass of the hydrogen atom.

736. Röntgen Rays. The negatively charged cathode ray particles traveling with such a high velocity must possess considerable energy. J. J. Thomson has shown mathematically that when an electrically charged particle is suddenly brought to rest an electromagnetic disturbance is produced in the surrounding medium and travels outward from the suddenly arrested particle. This condition is fulfilled when a cathode ray particle is suddenly brought to rest by striking against any solid body. In 1895 Röntgen observed that some sort of radiation was produced outside an ordinary cathode ray tube. Phosphorescent bodies placed near the tube were strongly affected and a photographic plate in the neighborhood became blackened. These radiations have been called Röntgen rays after their discoverer. The name first applied to them was X rays and this name is still often used.

The Röntgen rays, which originate at the point where the cathode rays strike a solid body, differ from cathode rays inasmuch as they are able to penetrate bodies of considerable thickness. Their penetrating power, as well as some of their other properties, depend upon the conditions existing within the tube from which they originate. With a very low pressure within the tube and, consequently, a large potential difference between the electrodes, the rays produced are very penetrating, being capable of going through several inches of wood and even several millimeters of lead. Such rays are usually called "hard rays." In the case of a higher pressure and smaller difference of potential the rays are less penetrating and are called "soft rays." Different substances absorb the rays of any particular type to a different degree. Generally speaking dense substances produce greater absorption.

It is this variation in the absorptive power of substances which enables us to make Röntgen ray photographs. Röntgen rays act upon a photographic plate in a manner similar to ordinary light and the effect produced depends upon the intensity of the rays. Thus a photograph of the bones of any portion of the human body may be obtained, for the bones being denser than the flesh absorb the rays more and, consequently, the intensity of the rays which have traversed the bones is less than the intensity of those which have passed through only the flesh.

The Röntgen rays travel in straight lines with very high velocity. Marx has shown recently that they travel with the velocity of light, that is, 3×10^{10} cms. per sec. No evidence has as yet been found of any diffraction of the rays when they pass from one medium to another, nor has it so far been possible to deflect the rays by a magnetic field.

737. Focus Tube. With a view to giving a clear idea of the method of producing Röntgen rays we will briefly describe the type of tube commonly used to generate them. There are various modifications, but they all conform to the same general principle which is represented diagrammatically in its simple form in Fig. 518.

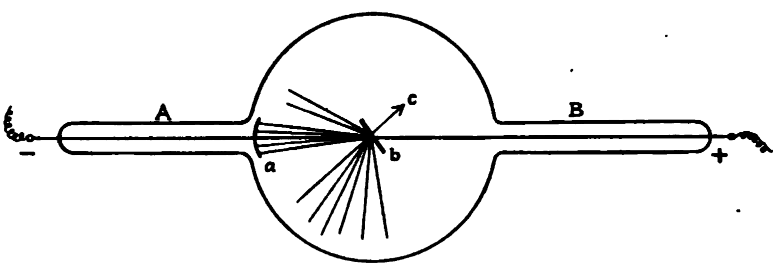


FIG. 518.

AB is a large glass bulb. The cathode *a* consists of a concave piece of metal, usually aluminum. The cathode rays proceed normally from the surface of *a* and on account of its concavity are brought to a focus at the point *c*, and hence the name focus tube. The anode *b* consists in its simplest form of a flat platinum plate which is placed at an angle of 45° to the axis of *a* and so that the center of *a* is at the point *c*. The Röntgen rays travel outwards in all directions from *b*. To generate the rays the electrodes are connected to the terminals of the secondary of an induction coil or to a Wimshurst machine.

738. Conductivity of Gases Produced by Röntgen Rays. Probably the most striking property of Röntgen rays is their power to cause gases to become conductors of electricity. If a well insulated body, such as the leaves of a gold leaf electroscope *E* (Fig. 519), be charged up in thoroughly dry air the charge will be retained for many hours. If, however, a beam of Röntgen rays pass through the gas surrounding the leaves they will immediately lose their charge and collapse, showing that the air must have become conducting, allowing the charge to leak away. (In the electroscope shown the leaves hang from a bead of sulphur, which is a good insulator, and may be charged through the bent wire shown. For quantitative work the leaves are observed through a reading telescope.)

Instead of the rays falling directly upon the gas surrounding

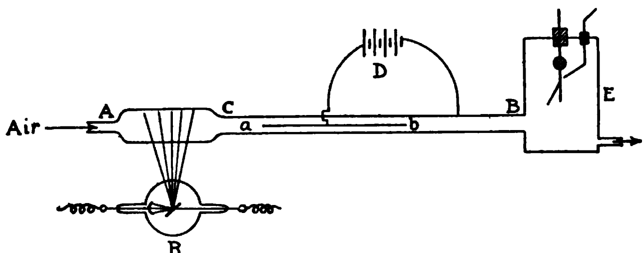


FIG. 519.

the leaves of the electroscope let a system be arranged as shown in Fig. 505.

AB is a metal tube through which a stream of air may be sent and which leads into an electroscope *E*. If the Röntgen rays fall upon the air in the part *AC* no effect is produced in the electroscope as long as there is no stream passing through the tube, but as soon as a stream of air is passed through the tube into *E* the leaves lose their charge. This conducting property imparted to the air by the rays, therefore, may be transported along with the air. If a plug of cotton wool be placed in the tube at *C*, or if the air be bubbled through water, after being acted upon by the rays this conductivity is entirely destroyed. If an insulated wire *ab* be introduced in the center

of the tube CB and a strong electric field be established between the wire and the tube, by connecting the wire to one pole of a battery and the tube to the other pole, the air loses its conductivity in passing through the tube.

It is also observed that this conductivity of the air persists for a short time after the rays have ceased. If the air is allowed to remain stationary for a few seconds after the rays have ceased and then drawn into the electroscope, without passing through an electric field, the leaves will be discharged.

The removal of this conducting power from the gas by filtering it through cotton wool or water indicates that the conductivity must be due to something mixed with the air, while its removal by an electric field shows that, whatever it may be that is mixed with the air, it must carry an electric charge.

The air rendered conducting by Röntgen rays discharges the leaves of an electroscope with equal facility whether they are charged positively or negatively.

Suppose again that A and B, Fig. 520, are two parallel metal

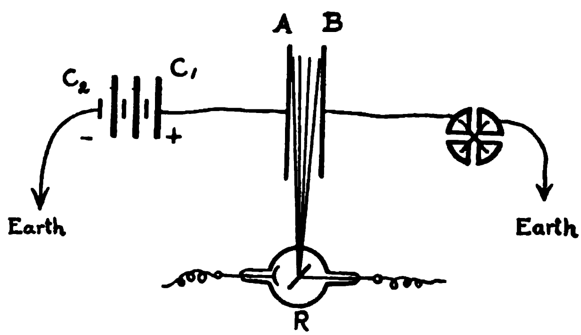


FIG. 520.

plates placed a few centimeters apart in air and let A be connected to one pole of a battery while the other pole is connected to earth; let B be connected to one pair of quadrants of a quadrant electrometer while the other pair of quadrants is connected to earth. If a beam of Röntgen rays be passed between these plates it will be observed that B immediately begins to receive a charge, as indicated by the deflection of the electrometer needle. It will continue to charge up as long as the rays are acting, but

will cease if the rays cease. If C_1 is the positive pole of the battery then B will receive a positive charge, but if the poles be reversed B will receive a negative charge. The rays thus apparently cause a transference of electricity through the air to B and the sign of the electric charge given to B depends upon the sign of A.

If the potential difference between A and B be altered the charge received by B in a given time alters, that is, the current between A and B, depends upon the voltage. The current through the gas does not, however, obey Ohm's Law, for if the current corresponding to different voltages be measured and a curve plotted showing the relation between current and voltage, it will assume the form shown in Fig. 521 instead of being a straight line. It will be seen that for small voltages the current obeys Ohm's Law, but it soon begins to fall off and finally reaches a constant value even for a large increase in voltage. This characteristic curve has been called a **saturation curve** on account of its similarity in form to the saturation curve in the magnetization of iron. The current corresponding to the flat part of the curve is called the **saturation current.**

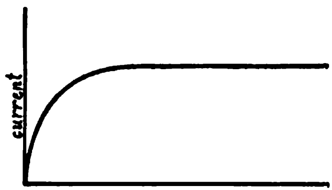


FIG. 521.

The current through a gas differs very markedly in another respect from the current through metals or liquids. When the distance between two electrodes immersed in a liquid is increased the current decreases on account of the increase of resistance between the electrodes, but in the case of a gas the saturation current increases when the distance between the plates is increased. Within certain distances the saturation current is proportional to the distance between the plates.

739. Theory of Ionization. These along with other properties of a gas rendered conducting by Röntgen rays led J. J. Thomson and E. Rutherford in 1896 to formulate the *ionization theory of gases,* which has now become firmly established by experiment. According to this theory the Röntgen rays, when they pass through a gas, cause the molecules of the gas to be broken up into positively and negatively charged carriers of electricity

called ions. This process of breaking up the molecule is called ionization and the gas is said to be ionized. From each molecule ionized two ions, having equal charges but of opposite sign, are produced. The transference of electricity through the gas is due to the movement of these charged carriers under the influence of an electric field. The positive ions are attracted to the negative electrode and the negative ions to the positive electrode, and the movement of these electric charges constitutes a current. When the gas is passed through the tube with a central wire, between which there is an electric field, the positive and negative ions are attracted to the negative and positive electrodes respectively and thus removed. When the gas is passed through cotton wool the ions are caught by the wool.

This theory also explains the saturation curve for a current between two plates. The greater the potential difference between the plates the greater the force pulling the ions out of the gas, and, consequently, the faster they will move towards the plates. The quicker they move the less chance there will be of their recombining, therefore, a greater number of ions would actually reach the plates per second: this number would be proportional to the voltage. The current, which is the quantity of electricity transferred to the plates per second, must be proportional to the number of ions reaching the plates in that time and, therefore, to the potential difference provided this is not too high. But when the voltage reaches a certain value the ions move so fast that they practically all reach the plates before they have time to recombine, and the current could not be increased further even by a higher voltage, as the number of ions removed could not be augmented.

The increase of current between two plates when the distance between them is lengthened is also easily explained by this theory. When the plates are placed farther apart the volume of gas acted on by the rays is increased and, consequently, the number of ions produced grows greater in the same proportion, and a greater number will reach the plates per second and the maximum current be raised.

740. Effect of Conditions on Ionization. The amount of ionization produced in a gas is dependent in a marked degree upon existing condi-

tions. The nature and quality of the ionizing rays determine the number of ions produced in any given gas. Cathode and Röntgen rays, for instance, differ in ionizing power and even Röntgen rays differ among themselves in this respect. Penetrating rays of any type are usually less powerful ionizers than those less penetrating.

For a constant ionizing source the number of ions produced in a given volume of gas is found experimentally to be directly proportional to the pressure. The number of molecules is also proportional to the pressure and, consequently, with increase of pressure there are more molecules to be ionized.

Temperature on the other hand has, as far as is known, no effect on ionization, if the density of the gas is kept constant. This question has been investigated by McClung over a considerable range of temperature for the ionization produced by X rays, and also by Rümelin for the rays from radium, and no effect has been observed.

The ionization is also dependent upon the nature of the gas. Heavy gases absorb the various types of radiations more than do the lighter gases and the greater the absorption the greater the ionization. In the case of some types of rays, such as the γ rays of radium, the ionization is approximately proportional to the density of the gas, but in the case of some of the less penetrating types of rays wide departures from this law occur.

741. Recombination of Ions. Another marked phenomenon is that when the rays begin to ionize the gas the ions gradually increase in number until a steady state is reached, when no further increase will take place no matter how long the rays act. As the rays are continually producing ions they must be disappearing at the same rate as they are being produced when this steady state is reached. Being positively and negatively charged bodies and being in motion they collide and neutralize each other electrically and disappear as far as producing any conductivity is concerned. The rate at which this recombination takes place is an important factor in determining the current through a gas.

742. Diffusion of Ions. The ions of an ionized gas are in motion and if there is an excess of ions in one part of the gas they will diffuse to the other part. If the ionized gas is in an enclosed vessel the ions will diffuse to the sides of the vessel and disappear from the gas. Sometimes, in a very confined space, the loss of ions by diffusion is even more important than the loss of recombination.

Townsend found that the rate of diffusion depended upon the nature of the gas, as was to be expected. The diffusion of the ions through the heavy gases is slower than through the lighter gases. He also found that in dry gases the negative ion always diffused faster than the positive, but if the gas contained considerable moisture the rates of diffusion of the positive and negative ions are much more nearly equal. This unequal

diffusion of the ions of opposite sign explains the phenomenon which is so often observed, that if an ionized gas containing equal numbers of positive and negative ions is passed through a metal tube it emerges positively charged. The negative ions diffusing faster to the sides of the tube than the positive ions, more of them are eliminated and hence the gas emerges with an excess of positive electricity.

The rate at which ions diffuse through gases is much slower than the rate of interdiffusion of ordinary gases. For instance, the rate for air and carbon dioxide is over five times as great as for the positive ion to diffuse through moist carbon dioxide. Heavy gases diffuse slower than light gases. The natural conclusion is that the mass of the ion in carbon dioxide is large compared with the mass of the molecule.

These facts have led to the theory that both the positive and negative ions, at ordinary pressures, consist of a cluster of molecules surrounding a charged nucleus. Ionization is considered to consist in separating a negative electron from the neutral molecule and then the electron becomes loaded with a cluster of molecules and forms the negative ion under ordinary conditions. The positive ion consists to begin with of the molecule deprived of the electron and then a cluster of molecules is formed about this positively charged center. This theory accounts for the fact that the positive and negative ions diffuse more nearly at the same rate in moist than in dry gases, for in dry gases the negative ion is smaller, but in a moist gas it becomes more loaded up with moisture than the positive ion and its rate of diffusion decreases more rapidly. This theory is supported by the fact that as the pressure of the gas is lowered the coefficient of diffusion of the negative ion increases faster than that of the positive. It has been shown by J. J. Thomson and by Townsend that at low pressures the negative ion is the same as the electron. These facts point to the conclusion that the negative ion at low pressures loses the cluster of molecules surrounding the electron.

743. Mobility of Ions. When an electric field is applied to an ionized gas the ions move under the influence of the field. It is of importance to consider the velocity with which they move. The velocity of the ions under a potential gradient of one volt per cm. is generally termed the mobility of the ions. The first experimental determination of this velocity was made by Rutherford. Later Zeleny made a series of very careful determinations of velocities and was the first to show that they were not the same for the positive and negative ions. He demonstrated that in any given gas the velocity of the negative ion is always greater than that of the positive. He also determined the absolute values of the velocities of both ions under different conditions.

The mobility of ions depends upon the gas in which they are produced, being greater in light than in heavy gases, and also upon the amount of moisture present. In dry air, for example, the velocities of the positive

and negative ions respectively are 1.36 and 1.87 cms. per sec. for a potential gradient of 1 volt per cm., while in hydrogen the corresponding values are 6.70 and 7.95 cms. per sec.

744. Ionization by Collision. In § 738 the current-voltage curve for a gas at atmospheric pressure showed a final maximum current. In a gas at low pressures, in the neighborhood of 1 mm. of mercury, a new phenomenon appears and the corresponding curve for current and voltage assumes the form shown in Fig. 522. For low voltages the part of the curve up to a point *A* is of the same form as the saturation curve at atmospheric pressure, but when the voltage is increased beyond a certain amount the current begins to increase again, at first slowly and then very rapidly. The increase of current beyond the point *A* must be caused by an increase in the number of ions due to some cause other than the original ionizing agency. This larger number of ions has been explained by the theory that if an ion is moving with sufficient velocity it will produce more ions by collision with the molecules of the gas. The cathode ray particle, or electron, is capable of producing ions when moving rapidly through the gas, for if the cathode ray stream be allowed to pass between two electrodes an ionization current may be observed. To ionize a molecule a certain amount of energy is required. A moving ion possesses kinetic energy and if its velocity is great enough it will possess sufficient energy to ionize a molecule with which it may collide. The kinetic energy depends upon the velocity and this, in turn, depends upon the electric field and upon the opportunity the ion has of acquiring speed among the molecules of the gas. At atmospheric pressure the molecules are so close together that the ion is not able, between two collisions, to acquire sufficient velocity in ordinary electric fields to ionize a molecule, but at low pressures the molecules are so few in number and so far apart that the ion may acquire suffi-

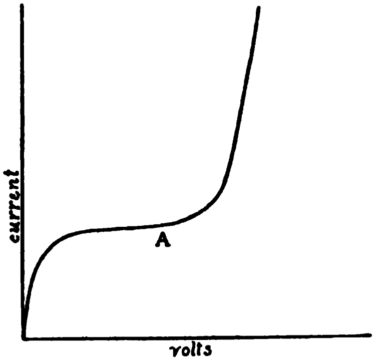


FIG. 522.

cient speed between collisions to ionize any molecule which it strikes. This production of ions by collision is only observed for ordinary electric fields at pressures below about 30 mm.

The above theory of ionization by collision furnishes a very satisfactory explanation of the electric spark through a gas at atmospheric pressure. There are always in gases a few ions which can be detected only by sensitive instruments. If a voltage high enough to produce a spark is established between two points, the few ions naturally present in the field will acquire a velocity sufficient to ionize any molecules against which they strike; these new ions will in turn produce more ions, and so the number will increase very rapidly until there are enough to carry a current, and this current is the electric spark.

745. Condensation of Water-vapor by Ions. It was known for some years that if dust particles were present in a damp gas the water vapor would condense around these nuclei when a sudden expansion of the gas took place. After the discovery of Röntgen rays, Richarz showed that if a beam of X rays were allowed to fall upon a steam jet, condensation took place. Some maintained that this effect was due to dust, but Richarz attributed it to the presence of ions. In 1897 and later C. T. R. Wilson made a very valuable series of experiments on this question and proved that ions do act as nuclei on which water vapor condenses when moist air is suddenly cooled by expansion.

Wilson also demonstrated that water vapor condenses more easily around the negative ions than around the positive ones. This is a confirmation of the theory that the cause of the greater diminution of the velocity and rate of diffusion of the negative ion in moist gases is due to the negative ion becoming more easily loaded with moisture than the positive one.

746. Charge Carried by an Ion. This property of ions to act as condensation nuclei has been utilized by J. J. Thomson to determine the absolute value of the charge carried by an ion. When an expansion takes place in ionized air water drops form around the ions and fall under the action of gravity. Sir George Stokes has shown that a drop of water of radius, r, falls through a gas of viscosity, μ, with the velocity, v, given by the equation

$$v = \frac{2}{9}\frac{gr^2}{\mu}$$

where g is the acceleration of gravity. The velocity, v, can be measured by observing the rate at which the cloud falls under the action of gravity, and since g and μ are known r may be determined. If m is the mass of water deposited and n the number of drops per c.c. then $m = n \times \frac{4}{3}\pi r^3$, since the density of water is unity. The amount of water vapor deposited when a known expansion occurs can be easily calculated from well known thermal considerations and, therefore, m may be determined. Knowing m and r the number of drops, n, which is the same as the number of ions, is easily calculated.

If all the ions present be extracted by an electric field between two electrodes in the usual way, the total charge carried by all the ions can be measured. Knowing, therefore, the number of ions and the total charge on them the charge carried by each one is determined.

By the latest determination which Professor Thomson has made he showed that

$$e = 3.4 \times 10^{-10} \text{ electrostatic units.}$$

He has also shown that the charge carried by the ions in hydrogen and oxygen has the same value and that it does not depend upon the source from which the ions are produced. These results seem to indicate that the charge carried by a gaseous ion is the same under all circumstances, and it appears that it might be taken as an invariable unit of electricity.

747. Ionization by Ultra-violet Light. If ultra-violet light rays fall upon the clean surface of a plate of zinc which is negatively charged the plate will lose its charge, while if the plate be uncharged to begin with it will acquire a positive charge. If the plate is positively charged to begin with no loss of charge takes place. These effects, which are called photo-electric effects, have been shown to be due to the liberation of negative corpuscles, or electrons, from the metal by the action of the ultra-violet light.

This **photo-electric effect** may be produced by allowing ultra-violet light to fall upon other metals besides zinc, such as sodium, potassium, lithium, etc. For this purpose the ultra-violet light may be obtained from an ordinary arc lamp, or from the spark of an induction coil between zinc, cadmium or iron terminals.

748. Ionization by Hot Metals. If a metal electrode be placed near to a metal wire and the latter be then heated until it begins to glow, a current through the gas will be produced and the electrode will receive a charge. The charge received by the electrode and the current depend upon several conditions. Temperature is an important factor. A platinum wire heated to redness will under some conditions give a positive charge to the other electrode, but if heated to white heat the charge is negative. The charge is also influenced by the pressure of the gas. The electrification also depends upon the nature of the gas around the wire and the material of the wire. The behavior of hot metals is somewhat irregular but in general metals and carbon heated to incandescence in high vacua give off negatively charged carriers. The ratio of the charge to the mass of these cariers has been shown to be the same as for the cathode ray particles and the electron liberated by ultra-violet light at low pressures. This along with other considerations has led to the theory that these negative corpuscles are distributed throughout the volume of metals at all temperatures, but when the metals are heated to incandescence the corpuscles then acquire sufficient energy to escape into the surrounding space.

749. Ionization by Flames. Gases around flames contain ions and conduct electricity. If two electrodes are placed some distance apart in an ordinary Bunsen flame quite an appreciable current is observed which may be measured by a galvanometer. If the air surrounding such a flame be drawn away from the flame it is found to be still a conductor. The ions which have been produced in the gas by the flame appear to be much larger than those produced in other ways, for their velocity has been measured and found to be much less than that of other ions. It is due to this conducting power of flames that when an insulator has received an electrostatic charge it may be discharged by simply passing a Bunsen flame over it.

RADIO-ACTIVITY.

750. Discovery of Radio-Activity. The phosphorescent action of Röntgen rays led physicists to investigate phosphorescent substances and Becquerel in 1896 found that the double sulphate of uranium and potassium emitted a radiation which produced an effect upon a photographic plate similar to that of X rays. He later examined other compounds of uranium as well as the element itself and found that they all possessed this power. Although the phosphorescent action of Röntgen rays pointed the way to this discovery it has since been shown that there is no connection between the rays emitted by uranium and its phosphorescence for some compounds which are not phosphorescent emit the rays.

Becquerel and other experimenters showed that these radiations from uranium were capable of discharging electrified bodies and later Rutherford showed that this power of discharging electrified bodies was due to the production by these rays of ions in the gas, similar to the ions produced by Röntgen rays.

If the rays from uranium or one of its compounds be allowed to pass between two parallel plates, between which there is a difference of potential, a current will pass through the air just

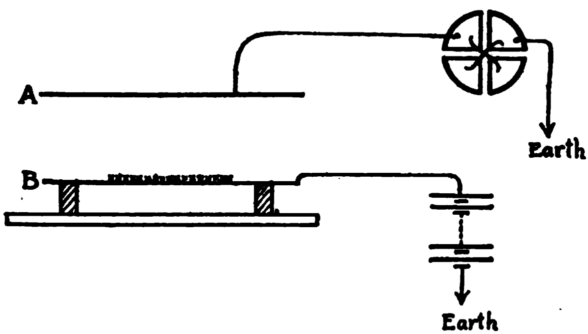


Fig. 523.

as in the case of a gas ionized by Röntgen rays. The current may be measured in a manner similar to that described in the previous section on ionization. The method may be represented diagrammatically in Fig. 523.

A and *B* are two insulated metal plates. The upper plate *A* is connected to one pair of quadrants of an electrometer, the other pair being to earth. If a layer of one of the compounds of uranium be sprinkled on the plate *B*, as indicated, an ionization current will be produced between *A* and *B*.

The results obtained by observations on uranium and its compounds indicate that this power of radiation belongs to the uranium itself and not to the substances with which it is associated, and that the radiations are emitted spontaneously without the aid of any outside agency. It has been shown that this property does not deteriorate with time. Uranium as well as other bodies, which we shall see later possess similar properties, are called ***radio-active bodies.***

751. Other Radio-Active Substances. The discovery of the radio-active property of uranium led to the examination of other substances. Schmidt, and independently Mme. Curie, discovered that the element thorium and its compounds possess radio-active properties. The photographic action of thorium was found to be distinctly weaker than that of uranium, while the ionizing action was about equal to that of uranium, but was very irregular. The cause of this irregularity will be discussed later. A very systematic examination of a large number of minerals containing uranium and thorium was then undertaken. Using the electrical method the current produced between two plates by a given amount of each of the minerals was measured. The results showed that all these minerals containing uranium or thorium were radio-active, but the important point observed was that several specimens of pitchblende, as well as some other minerals, were several times more active than uranium itself. It was at first thought that this abnormal activity of some of the minerals might be due to the particular chemical combination in which the uranium existed, but this was disproved by preparing one of the componds artificially, and it was found to possess only the normal amount of activity which would be expected from the amount of uranium it contained. This led to the conclusion that there must be some other and more active substance in pitchblende. M. and Mme. Curie then investigated this question chemically and discovered two new active bodies.

The first of these substances to be separated by purely chemical means was found to be very much more active than uranium and it was given the name polonium in honor of Mme. Curie's native country. Polonium differs from uranium in the essential particular that its activity is not constant but gradually dies away with time. In some cases it was found that at the end of about six months after preparation the activity had fallen to half its original value.

The other active substance discovered in pitchblende was found to be enormously more active than uranium. In its pure state it is about a million times more active, and it was called radium by the discoverers. Radium is probably the most remarkable and

interesting of all the radio-active substances, and by the study of its properties an enormous amount of information has been obtained in regard to the most remarkable processes going on in nature in connection with these radio-active bodies. The quantity of radium existing in pitchblende is almost infinitesimal, about a ton of pitchblende containing only a few milligrams of pure radium. Radium is found in varying quantities in a number of minerals and in various parts of the world, but the chief source at present known is in the pitchblende found in Bohemia.

In practice radium is not separated from its compound but is usually employed in the form of the bromide, and what is often called "pure radium" is really "pure radium bromide." It also forms other compounds, such as the chloride, sulphate, etc., and these salts are all naturally phosphorescent and their radiations produce phosphorescence in various substances such as platino-barium cyanide, willemite, etc.

Debierne, in analyzing residues from pitchblende, discovered a very active substance which he called actinium. The properties of actinium are very similar to those of thorium, but the former is very many times more active than the latter. Actinium besides being strongly radio-active is capable, like radium, of producing phosphorescence in such substances as zinc sulphide, willemite, etc.

752. Three Types of Rays. In the early examination of the radiations from uranium Rutherford showed that the rays emitted were of a complex nature. He found at first that there were two distinct types of rays, one type which were easily absorbed by solid bodies, and a second type which were more penetrating and, besides, could be easily deflected from their path by a magnetic field. The former he called α rays and the latter β rays. Later it was shown that there was still a third type emitted which were extremely penetrating and could not be deflected by a magnetic field. These were called γ rays. The four radio-active substances uranium, thorium, radium and actinium, under normal conditions, give out these three types of rays. Polonium, however, emits only α rays.

753. General Properties of Rays Emitted by Radio-Active Bodies.

The following distinct properties serve to differentiate the different type of rays. In general, the rays which produce the greatest photographic action produce the least ionization. Also, the more penetrating the rays the less efficient are they as ionizers. The α rays are very easily absorbed, being entirely absorbed by a sheet of aluminum of 0.1 mm. in thickness. A thickness of about 5 mm. of aluminum is capable of absorbing most of the β rays, but a very great thickness of metal is required to absorb the γ rays, as they will pass through several centimeters of lead.

The α rays are the most efficient ionizers. When the three types of rays are acting simultaneously on a gas by far the greater part of the ionization is produced by the α rays. The β rays produce more ionization than the γ rays.

On the other hand, the β rays are much more active in their action on a photographic plate than are the α rays. The photographic action of the γ rays is very small; in fact, in the case of uranium and thorium no photographic action of the γ rays has as yet been detected. In making experiments with these radiations this difference between the ionizing and photographic properties of the α and β rays especially, must be carefully taken into account, for otherwise contradictory results are apt to be obtained. Take a case when only one kind of rays is present, such as the α rays for instance. If the electrical or ionization method is used quite a large effect may be observed, while if the photographic method is employed the effect may be very small or may even not be detected at all.

The β rays are very easily deflected by a magnetic or electric field, but it is very difficult to bend the α rays, while a magnetic or electric field has no effect whatever upon the γ rays. The fact that the α and β rays are thus deviated shows that they must carry an electric charge. The α and β rays are bent in opposite directions by the same field and, consequently, they must be oppositely charged. The direction in which they are bent shows that the α rays are positively and the β rays negatively charged.

754. The β Rays. It was observed in 1899 and later by several experimenters that some of the rays emitted by radium compounds were easily deflected by a magnetic field, while the other rays

were apparently not deviable. The deviable rays were shown to be the same as the β rays of uranium, which were also shown later to be deviable by a magnetic field. Becquerel, using the photographic method, showed that the β rays of radium behaved in every respect like cathode rays. They, consequently, must be negatively charged particles moving with high velocity.

The fact that the β rays carry a negative charge has been proven by M. and Mme. Curie. They enclosed a metal plate in a solid insulator to prevent loss of charge by conduction through the gas. They allowed the β rays to pass through the insulator and fall upon the metal plate which was connected to an electrometer. When the β rays fell upon this plate it received a negative charge.

Since the β rays are negatively charged particles their deviability by a magnetic and electric field naturally follows. Becquerel has made use of this deviability to determine their velocity and the ratio of the charge to the mass. By allowing a narrow beam of the rays to pass between two parallel plates, between which there is a large difference of potential, or between the poles of a magnet, the deviation produced by the magnetic or electric field may be measured by observing the movement of the impression on a photographic plate. Combining these two deflections, in a manner somewhat similar in principle to that used in the case of the cathode rays described in the preceding chapter, Becquerel determined the velocity and the ratio of e/m for the β rays. He found the velocity to be about 1.6×10^{10} cms. per second. The velocity of cathode rays we have seen is about 2.8×10^{9} cms. per second, so the velocity of the β rays is considerably greater than that of the cathode rays. Using the same rays he found the value for e/m which does not differ much from the value found by J. J. Thomson for the cathode rays. He observed, however, that the rays did not all have the same velocity as some were bent more than others. He showed that the velocities varied from about 6×10^{9} to 2.8×10^{10} cms. per sec., the latter approaching very nearly the velocity of light which is 3×10^{10} cms. per sec. The β rays from radium appear, therefore, to be complex, being a mixture of rays of the same nature but travelling with different

speeds. The β rays from uranium differ from those of radium in this respect for the former appear to be homogeneous.

This complexity of the rays with regard to velocity led Kaufmann to examine whether the value of e/m for these rays varied with the speed. He showed that e/m decreased when the speed increased. Assuming that the charge on the β ray particle is constant the mass of the particle appears to increase with increase of velocity.

Several mathematical physicists have worked out from purely theoretical considerations that the apparent mass of a moving electron is due, either wholly or in part, to the electric charge in motion, that is when an electric charge is moving it appears to possess what corresponds to inertia, due to the fact of its being in motion. This apparent inertia according to this view is not due to material mass as we are accustomed to conceive of it, but is a result of the motion of the electric charge. These theoretical considerations further show that this apparent mass, which seems to be electrical in origin, increases with the speed of the moving charge. The experimental results of Kaufmann seem to confirm the theoretical view that the mass of the electron is due, wholly or in part, to the fact that the electric charge is in motion.

755. Magnetic and Electric Deviation of α Rays. The α rays were early distinguished from the β rays by the fact that they were much less penetrating than the β rays and also that they were apparently not deviable by a magnetic field. Their importance was not at first recognized and their true nature was not determined for some time after the nature of the β rays had been determined. As a result of some experiments on the relative ionization by α and β rays Strutt suggested that the α rays were positively charged particles emitted with great velocity. To test this the crucial experiment, of course, was to try to bend the rays by a magnetic field. Rutherford was the first to succeed in doing this.

His apparatus is shown diagrammatically in Fig. 524. Unless the radiation is very intense and the magnetic field very powerful it is extremely difficult to detect any deviation.

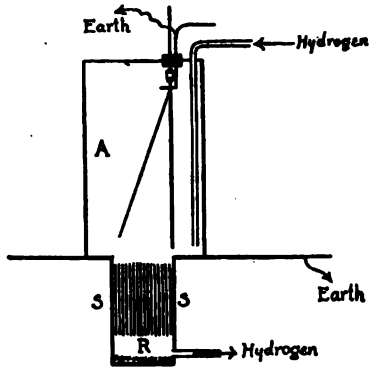


FIG. 524.

A is a gold leaf electroscope of the usual form. *SS* is a set of parallel brass plates separated by very narrow slits, the width of which was in some experiments as small as 0.042 cm. but varied for different experiments up to 0.1 cm. A quantity of radium, *R*, was placed below the slits and the rays passed up through them and into the electroscope where they ionized the air. Of course, the β and γ rays were also present but the ionization produced by the α rays was more than nine times that produced by the β and γ rays combined, so their presence did not affect the experiment. By applying a magnetic field in a direction parallel to the slits and at right angles to the plane of the paper the rays, if they are deviable, should be bent either to the right or left and strike the plates and be stopped before they could emerge beyond the slits. He found that by the application of the magnetic field over eight-ninths of the α radiation could be cut off, showing that the α rays could be deviated by the field. By a slight modification of the experiment he showed that they were bent in the opposite direction to that in which the β rays would be bent, indicating that the α rays must carry a positive charge.

Rutherford also succeeded in deflecting the α rays by an electric field using an apparatus similar to the one just described.

756. Velocity and Ratio e/m for α Rays. The deviability of the α rays by a magnetic and electric field made it possible to determine their velocity and the value of the ratio of the charge to the mass.

The latest results obtained by Rutherford and other experimenters show that within the limits of experimental error the value of e/m is the same for the α rays emitted by the various radio-active substances. The average experimental value obtained is about 5×10^3 electromagnetic units. Assuming that the charge on each particle is the same the mass of the α particles emitted by the different substances is constant.

Although the mass is constant yet the velocity of expulsion of the α particles is not the same for all substances, as it is found to vary from 1.56×10^9 to 2.25×10^9 cms. per second.

757. Mass and Nature of *α* Particle. These results enable us to obtain a more definite idea of the mass and nature of the *α* particle. The value of e/m for the atom of hydrogen liberated in the electrolysis of water is 10^4 electromagnetic units, while we have just seen that for the *α* particle e/m is 5×10^3. Recent experiments by Rutherford show that within the limits of experimental error the charge carried by the *α* particle is twice the charge carried by a gaseous ion and consequently twice the charge on the electrolytic hydrogen ion or atom. It follows from this that the mass of the *α* particle must be four times the mass of the hydrogen atom. Different hypotheses have been suggested as to what the *α* particle really is. Since it is atomic in size and of the same order as the atom of helium (whose atomic mass is 3.96 in terms of hydrogen) and since there does not seem to be any place according to the periodic law among the elements for a new one in that part of the series the most natural hypothesis is that it is an atom of helium carrying twice the ionic charge of hydrogen. Recent experimental evidence by Rutherford points very strongly to the conclusion that this is the correct hypothesis. This is supported by the experimental fact that helium is continually produced by both radium and actinium and also by the fact that helium is commonly found along with old radio-active minerals.

758. Absorption of *α* Rays. A distinguishing characteristic of the *α* rays is that they are very easily absorbed when passing through either gases or solids. The proportion of the rays absorbed by a given thickness of any solid may be determined by first measuring the saturation current produced by the rays, and then covering the radiating material with the absorbing solid and again measuring the current produced by the rays after they have passed through the solid. The absorbing layer must be very thin or else all the rays will be stopped. The most penetrating *α* rays known are completely absorbed by a thickness of only about 0.006 cm. of aluminum. The penetrating power of the *α* rays varies greatly with the different substances from which they are emitted.

The *α* rays are very easily stopped by gases, a few centimeters of air at atmospheric pressure being sufficient to absorb them, consequently, the ionization produced by them exists only within a few centimeters of the source from which the rays come. The absorption by gases depends upon the density, being in some cases proportional thereto, but not so in all cases. The absorption of the rays by gases is important as the degree of ionization produced by the rays depends upon the amount of the rays

absorbed, the relative ionization by the α rays in gases being directly proportional to the relative absorption.

759. The γ Rays. The third distinct type of rays given out by some of the radio-active substances differs very essentially from the α and β rays. The γ rays are extremely penetrating, being capable of passing through large thicknesses of solid matter. For instance, the γ rays given out by very strong radium bromide can be detected after passing through 30 cms. of iron. They are very much more penetrating than the X rays from a very hard X ray bulb. They ionize gases but to a very much less extent than either the α or β rays, the ionization being very approximately proportional to the density of the gas.

No one has as yet succeeded in deviating the γ rays by either an electric or magnetic field. They do not appear to carry any electric charge. It has, therefore, not been quite so easy to determine the real nature of the γ rays in a direct manner. Their great penetrating power and their non-deviability show a close resemblance to very hard X rays. The more penetrating X rays become, the more nearly does the conductivity produced by them become proportional to the density of the gas, which shows a resemblance between the very hard X rays and the γ rays.

We know also that X rays are produced by the sudden stopping of a moving electron, and it is reasonable to suppose that they would be produced by the sudden starting of an electron. Now experiment has shown that γ rays always occur in conjunction with β rays and the β rays we know are electrons. Consequently, it is reasonable to suppose that the γ rays are electromagnetic pulses produced by the sudden emission of the β particle, or electron, from the radio-active substance. The theory that the γ rays are similar in nature to very hard X rays, but of a more penetrating type, seems to be supported strongly by the evidence at present available, although it is very difficult to settle the question definitely by direct proof. Bragg has recently put forward a theory which he considers to be strongly supported by experiment to the effect that the γ rays consist of neutral pairs of positively and negatively charged particles.

760. Production of Uranium X and Thorium X. Crookes in 1900 showed that by a simple chemical process he could separate from

uranium a constituent which was many times more active photographically than the uranium from which it was separated and, in addition, the separation of this constituent left the uranium photographically inactive. This new and unknown constituent he called Uranium X, or Ur. X. Becquerel obtained similar results using a slightly different chemical process, and, on testing about a year later the Ur. X and the uranium from which it had been separated, discovered in addition the curious fact that the uranium had completely recovered its usual amount of activity while the Ur. X had entirely lost its activity. Rutherford and Soddy later succeeded in performing a similar chemical operation with thorium, separating a very active constituent from thorium, which they called Thorium X or Th. X and which acted in a manner very similar to Ur. X.

These phenomena have been thoroughly examined, both by the photographic and electrical methods, and it has been found in the case of uranium that after separation the Ur. X was very active

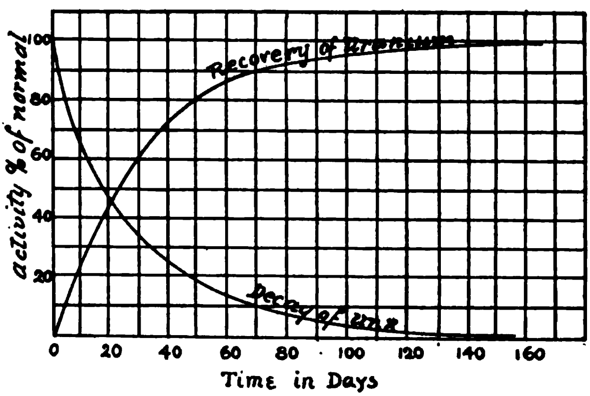


FIG. 525. (After Rutherford, Radio-activity.)

photographically but inactive electrically, because it gave out β rays but no α rays, while the uranium from which it had been separated was inactive photographically but still active electrically, due to the fact that it gave out α rays but practically no β rays. The Ur. X gradually lost its activity, while the uranium regained its β ray activity again, and the loss in the one instance and the

recovery in the other took place at the same rate. When the Ur. X had lost half its activity the uranium had regained half its original activity and each process took about 22 days. The way in which this occurred is shown very clearly by the curves in Fig. 511 which represent the activity of each at different times after separation, the ordinates representing activity and the abscissæ time in days. Similar results but of a slightly more complicated nature were observed for thorium, but the time taken for the activity of Th. X to decay to half its maximum value and that of the therium to regain half its activity was found to be only 4 days.

These results indicate that some process must be continually going on in these substances. Since the Ur. X which gives out β rays can be separated from the normal uranium leaving it devoid of β rays, therefore, the β rays must arise from the Ur. X, and since the uranium regains the β ray activity after separation more Ur. X must be formed in the uranium compound to give rise to these rays. This can be shown to be true, for Ur. X can be separated a second time after recovery has taken place. The activity of normal uranium does not change, consequently, there must be a state of equilibrium in the uranium in which Ur. X is being formed at the same rate as it dies away, so that the resultant activity remains constant. This is borne out by the fact that the rate of decay of Ur. X is the same as the rate of recovery of the uranium from which it was separated. Processes of a similar nature have been shown to be continually taking place in radium and actinium compounds.

These facts, along with a great deal of additional evidence, some of which we shall consider later, led Rutherford and Soddy to formulate the theory of successive changes in radio-active substances. According to this theory the different radio-active substances are gradually undergoing a process of transformation by which they are changing in regular succession from one product to another without the help of any outside agency. We shall see later that Th. X, for instance, is not lost when its activity completely decays, but it disappears as Th. X and changes into another product or substance. Most of these products give out radiations similar to those we have considered, but some do not give out any

at all and are, consequently, called rayless products. This is called the theory of successive changes and the different products into which each one is gradually transformed are called transformation products. The rates at which these changes take place vary very greatly for the different products, some changes only taking a few seconds to complete, while others extend over hundreds of years. The time it takes any one of these changes to be *half* completed is generally spoken of as the *period* of that transformation, as this time is usually much more easily determined experimentally with accuracy than the time of the complete change.

Actinium possesses a corresponding active constituent called Actinium X with properties similar to Th. X. As yet, however, no corresponding product has been discovered in connection with radium.

761. Emanations from Radio-Active Bodies. The early experimenters on thorium observed that the radiations given out by thorium compounds were very irregular. Rutherford investigated this irregularity and found that it was due to the emission of some sort of radio-active particles from the thorium compound. To these particles he gave the name "**emanation,**" and he found that it was not like the radiations which we have already considered, but acted in all respects like a gas. It will diffuse through porous solids and through gases and it may be carried away by a current of air. It is capable of ionizing a gas itself and of acting on a photographic plate. It does not itself consist of ions but has the power of producing ions in the gas, for it may be passed through cotton wool or bubbled through solutions without losing its power of ionizing a gas. This differs from a gas ionized in the ordinary way, for the gas will lose its ions under these circumstances while the emanation does not.

The emanation is not affected by an electric field. The electric field removes the ions produced by it but does not remove the emanation itself. The emanation cannot, therefore, consist of charged particles like the ions.

Both radium and actinium compounds give out an emanation possessing properties similar to the thorium emanation, but as far as is known at present uranium compounds do not give off any emanation.

Rutherford and Soddy investigated the effect of different physical and chemical agencies on the emanations from thorium and radium, and found that these emanations were chemically inactive, not being affected by the strongest reagents. They were not altered by being passed through a platinum tube raised to a white heat, nor by being cooled to the temperature of solid carbon dioxide. They found, however, that the emanation could be condensed when passed through a tube immersed in liquid air. This was a very important and crucial experiment, proving conclusively the gaseous nature of the emanation. They showed that the radium emanation condensed at about $-150°$ C. while the thorium emanation began to condense at $-120°$ C. but was not all condensed till about $-150°$ C. was reached.

It has very recently been shown that actinium emanation may be condensed under the same conditions and at the same temperatures as thorium emanations.

If the emanation be removed from the thorium, by drawing off into another vessel both it and the air with which it is mixed, its activity dies away very rapidly with time. Also if a quantity of thorium be placed in a closed vessel and the ionization current measured immediately and at short intervals it is found to gradually rise and finally reach a steady state. The rate at which the current rises in the closed vessed is exactly the same as the rate at which the separated emanation dies away. We have here a state of things similar to the case of thorium and Th. X where the activity of one rises at the same rate as the other dies away. An equilibrium state is reached when the emanation is produced as fast as it dies away. The period of the emanation, that is, the time taken for its activity to fall to half value, is about 54 seconds. The period for the radium emanation is much longer, being about 3.7 days.

The emanation is not produced directly by the thorium but is a direct product of Thorium X. Rutherford and Soddy have shown that when the Th. X is separated from the thorium the latter does not give off any emanation but gradually regains its emanating power. The separated Th. X, however, possesses strong emanating power but gradually loses it. These processes take place at exactly the same rate as the loss and regain of activity

by the Th. X and thorium respectively, which we have already considered. This accounts for the decay of the Th. X as it is continually changing into emanation. The emanation and Th. X are distinct substances having distinct properties.

762. Excited Activity. If a solid body be exposed in a closed vessel to the emanations from radium, thorium or actinium its surface becomes coated with an extremely thin solid deposit of very radio-active material. This active deposit is invisible, even under a microscope, but can be dissolved by certain acids and when the solvent is evaporated again it is left behind. It emits radiations which affect a photographic plate and ionize a gas. If a negatively charged wire be placed in a closed vessel containing the emanation the active deposit is all concentrated on this wire instead of being distributed on the interior of the vessel. By this means a very small wire may be made intensely radioactive.

The active deposit can be removed from a wire by rubbing with sand paper, but the quantity deposited is so extremely small that no increase in weight can be detected in a wire which has received an active deposit. This active deposit is not due to any action of the radiations given out by the radio-active compound but is a direct result of the presence of the emanation, for when no emanation is present no active deposit is observed and, in addition, the amount of excited activity is always proportional to the amount of emanation present.

If the negatively charged wire be exposed to the emanation for several hours and then removed and its activity tested at intervals, it is found to gradually die away with time, according to a law exactly similar to that for the decay of the emanations. The excited activity from thorium decays to half value in about 11 hours. It requires time for the excited activity to be deposited on the wire and the deposit increases until it reaches a maximum. This rate of increase is the same as the rate of decrease of activity when the wire is removed from the emanation. There must, consequently, be an exactly similar process going on here as we observed in connection with thorium and Th. X, and with Th. X and the emanation. Just as Th. X is continuously changing into the emanation the emanation is gradually changing into the active deposit and this in turn must be changing into something else.

If the wire be exposed to the thorium emanation for only a few minutes instead of several hours a different phenomenon is observed after removal of the wire. Instead of beginning to decay immediately after removal the activity, which at first is very small, gradually increases until it reaches a maximum in about four hours, and then it decays again at just the same rate as the activity for a long exposure decayed. When the exposure is a long one no initial increase is observed. Rutherford was the first to offer a satisfactory explanation of this phenomenon and he did so by supposing that the active deposit, instead of being one substance, is really made up of two distinct substances one of which is changing into the other. He called these two substances thorium A and thorium B and supposes that thorium A arises from the emanation and is deposited on the wire and then changes into thorium B, and then the thorium B changes into something else. For a short exposure the deposit will consist almost entirely of thorium A, as very little has had time to change into thorium B, and if we suppose that thorium A either gives out no rays at all or rays which produce a very small amount of ionization compared with those from thorium B, then the activity at first will be very small, due almost entirely to the very small portion of thorium B. As thorium A changes into thorium B the activity will increase until the change of A into B just balances the decay of B. Then, as more atoms of B will change per second than are produced from A, the activity will gradually decay. In the case of the long exposure this maximum has been reached before the wire is removed and tested and, consequently, the initial rise is not observed. It was thought for some time that thorium A gave out no rays but it has been shown recently that it emits a slow type of β rays. These two substances have distinct periods of decay which have been determined partly by experiment and partly by theoretical considerations, which space will not allow us to enter into here. Thorium A has a period of 11 hours and thorium B a period of 55 minutes. Hahn has more recently shown that thorium B is not a single substance but is complex, consisting of two distinct substances which he has called thorium B and thorium C, the former giving out α rays and changing into the latter, which gives out α rays and probably also β and γ rays.

The active deposit from the actinium emanation is very similar

to the active deposit from thorium, consisting of actinium A, actinium B and actinium C.

An examination of the active deposit from radium shows that the transformations taking place are more complicated than those of thorium and actinium. The decay curves when measured by the α rays are quite different from those obtained by the β or γ rays. The two latter give identical curves showing that the β and γ rays occur together. By a process of analysis similar to that used for thorium it has been shown that the active deposit from the radium emanation consists in the first instance of three distinct substances which have been called radium A, radium B and radium C. Radium A gives out only α rays; radium B was at first thought to give out no rays but it has recently been shown to give out β and γ rays, while radium C gives out all three types. The periods of these three products are 3, 26 and 19 minutes respectively.

It has been observed that after the greater portion of the excited activity of the deposit from radium has decayed, which takes place, as we have seen, in a few minutes, there is a small residual activity remaining which decays extremely slowly. This residual activity has been shown to consist of four distinct substances which have been named radium D, radium E, radium F, and radium G. The periods of transformation of these products are much longer than those of radium A, B and C. Radium G has been shown to be identical with polonium.

763. Radiothorium and Radioactinium. Hahn has recently shown that there could be separated from thorium by chemical means an intensely active substance which he called radiothorium, possessing all the characteristic properties of thorium, but many thousand times more active. It is from this substance that Th. X arises. He has also succeeded in separating still another substance called mesothorium which has later been shown to be itself complex, consisting of two products for which the names thorium 1 and thorium 2 have been suggested. These are intermediate products between thorium and radiothorium.

He has also shown that in the actinium series there is a product similar in properties to radiothorium which he has called radioactinium. It is produced from actinium and is in turn the parent of actinium X.

764. Heat Emitted by Radium and Thorium. Curie and Laborde discovered that radium is always hotter than its surroundings and emits heat at the rate of 100 calories per gram per hour. It has also been found that thorium acts similarly though in a minor degree. This is readily explained by the high velocity and kinetic energy of the α particles (§ 756) and the readiness with which they are absorbed (§ 758). Many of the particles that start within the radioactive body are absorbed by the body itself and their kinetic energy is transformed into heat.

765. Theory of Radio-Active Changes. We have seen that in the radio-active bodies continuous changes from one substance to another are taking place which so far have never been observed in any other class of materials. Each of these substances is entirely distinct from the others and has distinct physical and chemical properties. They, however, gradually decay and each one has a distinct and definite period of decay which distinguishes it from all the others. How do these changes come about? The disintegration theory or theory of successive changes furnishes the now generally accepted explanation.

According to the theory of J. J. Thomson atoms may be considered complex structures consisting of systems of positively and negatively charged particles in very rapid rotation and held together by their mutual forces in equilibrium. According to the disintegration theory this complex structure constituting the atom of radium (which we shall take as a typical example) becomes by some means unstable and one of the positively charged α particles is suddenly expelled with great velocity. The structure of the atom which remains is now different and constitutes the atom of a new substance, namely, the emanation. The atoms of the emanation are unstable and gradually change by the expulsion of another α particle, leaving a new structure, namely, the atom of radium A, and the process is continued throughout the successive changes. The processes are not identical in all instances, for in some cases an α particle alone is expelled, but in others β particles are expelled accompanied by γ rays, while in others all three types are given out.

Why do these atoms suddenly become unstable and break up without any apparent cause? Several explanations have been

offered to account for this, but the most probable one seems to be that if this system of charged particles, of which the atom probably consists, is in rapid rotation it must be radiating energy, and when sufficient energy has been radiated the mutual forces of the system no longer balance and one or more of the particles escape and cause disintegration. These atoms have an independent existence and distinct physical and chemical properties, but they differ from the atoms of ordinary non-radio-active elements in the fact that they are not permanent. To distinguish them from ordinary atoms the term metabolon has been suggested as a suitable name.

A few of these transformation products do not emit any rays at all and the change from them into the succeeding substance apparently takes place without the expulsion of any particles. These so-called rayless changes may be explained in either of two ways. The new product may be formed in this case simply by a rearrangement of the system of charged particles, but not with sufficient violence to expel any of the system, or it may be produced by the expulsion of one or more particles, but with a velocity too slow to ionize the gas. It has been shown that when the velocity of the α particle falls below 10^9 cms. per second it ceases to ionize the gas, and consequently an α particle expelled with a velocity below this minimum would escape detection since no ions would be produced.

This latter hypothesis suggests that all matter may possibly be undergoing a slow change in a similar manner, and that the only reason this change has been observed only in the so-called radio-active bodies and not in other non-radio-active bodies is that in the case of the radio-active bodies the charged particles are expelled with sufficient violence to ionize the gas while in other bodies they may be expelled but not with sufficient velocity to produce ions.

766. Radio-Active Elements. The following table contains a summary of all the active products at present known. On account of the incomplete state of the subject this list will no doubt shortly undergo changes by further discovery.

Radio-active Products.	Transformation Period.	Nature of Rays Emitted.
Uranium	5×10^9 years	α
↓ Uranium X	22 days	β and γ
↓ Ionium	?	α
↓ Radium	2,000 years	α
↓ Emanation	3.75 days	α
↓ Radium A	3 minutes	α
↓ Radium B	26 minutes	β and γ
↓ Radium C	19 minutes	α, β and γ
↓ Radium D (Radio-lead)	40 years	No rays
↓ Radium E	6 days	No rays
↓ Radium F	4.5 days	β and γ
↓ Radium G (Polonium)	140 days	α
↓ ?		

Radio-active Products.	Transformation Period.	Nature of Rays Emitted.
Thorium	10^{10} years	α
↓ Thorium 1 } Mesothorium	5.5 years	No rays
↓ Thorium 2 } Mesothorium	6.2 hours	β and γ
↓ Radiothorium	800 days	α
↓ Thorium X	3.7 days	α
↓ Emanation	54 seconds	α
↓ Thorium A	11 hours	β (slow)
↓ Thorium B	1 hour	α
↓ Thorium C	?	α, β and γ
↓ ?		
Actinium	?	No rays
↓ Radioactinium	19.5 days	α
↓ Actinium X	10 days	α
↓ Emanation	3.7 seconds	α
↓ Actinium A	36 minutes	β
↓ Actinium B	2.15 minutes	α
↓ Actinium C	5.1 minutes	β and γ
↓ ?		

REFERENCES.

THOMSON'S *Conduction of Electricity through Gases.*
STRUTT'S *The Becquerel Rays and Other Properties of Radium.*
RUTHERFORD'S *Radio-activity.*
RUTHERFORD'S *Radio-active Transformations.*

INDEX OF NAMES

Numbers Refer to Pages

INDEX OF SUBJECTS

Numbers Refer to Pages